이학박사 학위논문

Development of a Protein-Ligand Docking Program Based on Global Optimization

광역최적화에 기반한 단백질-리간드 도킹 프로그램 개발

2014년 2월

서울대학교 대학원
화학부 물리화학 전공
신 웅 희

Development of a Protein-Ligand Docking Program Based on Global Optimization

지도교수 석 차 옥

이 논문을 이학박사 학위논문으로 제출함
2014년 2월

서울대학교 대학원
화학부 물리화학 전공
신 웅 희

신웅희의 박사학위논문을 인준함
2013년 11월

위 원 장	신 석 민	(인)
부위원장	석 차 옥	(인)
위 원	이 상 엽	(인)
위 원	박 황 서	(인)
위 원	정 연 준	(인)

Abstract

Development of a Protein-Ligand Docking Program Based on Global Optimization

Woong-Hee Shin

Department of Chemistry

The Graduate School

Seoul National University

Protein-ligand docking has become an essential tool for computer-aided drug discovery since docking programs were first developed in 1980's. The goals of docking are to predict 1) the binding mode and 2) the binding affinity of a given protein-ligand complex accurately. Accurate prediction of binding mode requires appropriate sampling of both protein and ligand conformations. Many available docking programs sample ligand structures successfully because ligand has a relatively small number of degrees of freedom. However,

a lot of current docking programs treat receptor as a rigid molecule although receptor often adapts its shape to bound ligand because treating receptor flexibility is a very complicated problem. First of all, the large conformational space of receptor is a challenge for typical sampling methods. In addition, current energy functions such as empirical docking score functions or force field-based energy functions do not accurately describe flexible receptor-flexible ligand interactions yet.

In this thesis, the development process of an efficient docking program that treats receptor flexible, called GalaxyDock, is described. A powerful global optimization technique, called conformational space annealing, was employed for simultaneous sampling of the conformational space of protein and ligand. In addition, a new energy function for flexible-receptor docking was designed by combining the AutoDock energy function and a knowledge-based ROTA potential. With these components for sampling and scoring, GalaxyDock shows high performances in the binding pose prediction and virtual screening benchmark tests when compared to other state-of-art docking programs. This result suggests that the GalaxyDock program can provide a firm basis for further method developments and for practical applications to *in sillico* drug discovery processes.

keywords: protein-ligand docking, global optimization, virtual screening, computer-aided drug discovery

Student Number: 2008-20327

Contents

List of Figures

List of Tables

1. Introduction

1.1. Overview of protein-ligand docking

Protein-ligand docking (PLD) is a method that predicts 1) in which pose a given ligand binds to a protein's binding pocket and 2) how strongly the ligand binds to the protein. The first docking program was developed in the early 1980's (Kuntz *et al.*, 1982). Since then, a number of docking programs have been developed (Morris *et al.*, 1998, Zsoldos *et al.*, 2007, Rarey *et al.*, 1996). Owing to its powerful prediction ability, many successful applications to drug discovery processes targeting proteins with available structures have been reported (Lavecchia *et al.*, 2006, Rogers *et al.*, 2006, Venkatesan *et al.*, 2010)

Similar to many other modeling problems, PLD has also two importantly components: 1) sampling candidate binding poses and 2) evaluating the sampled poses. The first popular program of PLD is UCSF DOCK developed by Kuntz (Kuntz *et al.*, 1982) as mentioned above. It evaluates binding affinity of candidate poses by calculating and adding van der Waals potential and Coulomb potential. It incorporates Dead-end elimination (DEE) to search binding modes of ligand. It assumes binding of protein-ligand as a lock-and-key model. In the following two sections, sampling methods and scoring functions that are implemented in PLD programs will be reviewed.

Another important issue in PLD field is treating protein flexibility. In nature, receptor changes its shape when partner binds.

This phenomena could be explained as 'induced-fit' model or 'conformational selection' model. However, in the early era of PLD field, protein flexibility was not be able to treated because the degrees of freedom (DoF) increases. Nowadays, as the computing power increases, those increament of DoF could be treated efficiently. Treatment of receptor flexiblity will also be covered in section 1.4.

1.2. Sampling methods of protein-ligand docking

Sampling method of PLD program generates candidate binding pose of ligand in receptor binding pocket. PLD programs can be classified into three groups by sampling method. First group of them uses Dead-end elimination (DEE). The popular programs of this group are FlexX and UCSF DOCK (Kuntz *et al*., 1982, Rarey *et al*., 1996). FlexX divides ligand into several fragments and receptor binding pocket is described by interaction points. The three of ligand fragments and three interaction points of receptor generate triangle. Those two triangles are matched and compared their shape and chemical complementarity. If they are matched better than other combinations, the fragments of triangle is located to the surface of those interaction points. The remaining fragments of ligand will be placed considering geometry. The pose generation of newly added fragments of ligand is governed by DEE.

Second group uses single trajectory by Monte Carlo (MC)

methods. One of the popular program among the group is RosettaLigand (Davis and Baker, 2009). The variation of MC such as Tabu search is also used.

The last group uses population-based optimization methods such as genetic algorithm (Morris *et al*., 1998, Jones *et al*., 1995). Almost of docking programs are included in this group. The population-based method starts with a number of population that each member of them contains information of ligand conformation. Then those population searches the lower energy region heuristically. Recently, more efficient global optimization algorithms such as ant colony optimization (Korb *et al*., 2009) and particle swarm optimization (Chen *et al*., 2007) have also been adopted to solve the sampling problem involving a large number of degrees of freedom.

1.3. Scoring problems of protein-ligand docking

The second component, scoring function is used to calculate the effective free energy of the state corresponding to candidate ligand binding modes. A scoring function is used to predict the binding mode with the lowest free energy and also to predict which ligand binds to a given receptor most strongly in a compound library. Docking scoring functions can be classified into three types (Huang *et al*., 2010). The first type is force field-based scoring functions. Such scoring functions are based on physical atomic interaction energies, adopting the

functional forms and parameters of available force fields like AMBER. In spite of its physical basis, such scoring functions have difficulties in effectively treating the solvation effect. Popular examples of force-field based scoring functions are DOCK (Meng *et al*., 1992) and AutoDock (Morris *et al*., 1998).

The second type of scoring functions is empirical scoring functions. Such scoring functions estimate the binding free energy by a weighted summation of empirical energy functions. The components of empirical functions can be any types of interactions in any functional forms. The functional forms are often very simplified ones such as piecewise-linear or quadratic functions. Typical energy terms include hydrogen bonding energy, hydrophobic interaction energy, desolvation free energy, and ligand conformational entropy. The relative weights of the constituting terms are determined by fitting to experimentally determined binding affinities. X-Score is one of the most popular empirical scoring functions that predict the binding affinity most accurately (Wang *et al*., 2002).

The last type is knowledge-based scoring functions. Such scoring functions are derived from experimentally determined structures. A principal assumption behind the derivation of knowledge-based scoring functions is that the frequency of observed interaction pairs in the structure database follows Boltzmann distribution. DrugScore is one of representative examples of knowledge-based scoring functions (Velec *et al*., 2005).

Although the two components of docking, sampling methods

and scoring function, were introduced independently above, they are highly inter-winded together in docking procedure. For example, efficient sampling of bound conformations driven by an accurate binding free energy landscape is the most desirable scenario in docking simulations.

1.4. Flexible-receptor docking

One of the most important contemporary issues in the PLD problem is how to incorporate flexibility of receptor. Many docking programs treat receptor as a rigid molecule, following the concept of the 'lock-and-key' model. However, receptor often changes its conformation when ligand binds as described by the 'induced-fit' model. Many failures in binding mode prediction and virtual screening are caused by improper treatment of receptor flexibility (Sousa *et al.*, 2006). Since the late 1990's, a number of flexible-receptor docking programs have been developed, and they can be classified into three categories (Shin and Seok, 2012).

The first category is ensemble docking. In ensemble docking, a pre-sampled ensemble of receptor structures represent receptor flexibility, and docking is performed on the ensemble structures. The receptor structures can be obtained from multiple crystal structures bound to different ligands (Bottegoni *et al.*, 2009), snapshots of molecular dynamics simulation trajectories (Wong *et al.*, 2005), conformations with

different side chain rotamer combinations (Clauβen *et al*., 2001), or structures generated from normal mode analysis (Cavasotto *et al*., 2005). The advantage of this kind of method is that the conformational space of protein is greatly reduced, and computation is very easy. However, the performance sensitively depends on the choice of the conformation ensemble.

The second category of flexible-receptor methods employs 'enlarged binding pocket'. In these methods, the binding pocket is effectively enlarged by softening the van der Waals potential (Sherman *et al*., 2006) or by alanine mutation of pre-selected side-chains (Bottegoni *et al*., 2008). After placing a ligand in the enlarged pocket, the resulting protein-ligand complex has to be refined with a harder potential or the side chains have to be recovered. This kind of methods can reduce the energy barriers by smoothing the energy landscape, enabling efficient sampling. However, such methods might produce false positives and unphysical binding poses. Therefore, a refinement step or a curation step are necessary (Sherman *et al*., 2006)

The last category of flexible-receptor docking methods employs simultaneous sampling of receptor and ligand conformations. Prior knowledge on the variable regions in the receptor can reduce the conformational search space. Advantage of this category of methods is that it does not require pre- or post-sampling of receptor conformations. However, simultaneous sampling requires increased computational time. A number of optimization techniques have been applied to simultaneous sampling for example by using Monte Carlo with minimization (Davis

and Baker, 2009) or genetic algorithm (Zhao and Sanner, 2007, Corbeil *et al*., 2007, Morris *et al*., 2009)

In all the three categories of the flexible-receptor docking mentioned above, the sampling and scoring problems become more serious than in the rigid-receptor docking. Considering flexibility of receptor enlarges the conformation search space, especially in the case of the simultaneous sampling method. In addition, proper energy function that describes the protein internal energy have to be included to score different protein conformations.

1.3. Outline of this thesis

In this thesis, the development procedure of a new protein-ligand docking method, called GalaxyDock, is introduced. GalaxyDock incorporates conformational space annealing (CSA) (Lee *et al*., 1997), a powerful global optimization technique, and a new scoring function (Shin and Seok, 2012) that combines AutoDock scoring function (Morris *et al*., 1998) to calculate protein-ligand interactions and ROTA scoring function (Hartmann *et al*., 2007) to calculate the internal energy of protein. A separate function that predicts the binding affinity is also introduced and tested. GalaxyDock has been tested on several benchmark test sets including those for rigid-receptor docking, flexible-receptor docking, and virtual screening. Comparisons with other state-of-art docking programs were also performed.

The following sections of this section cover the progress of the GalaxyDock development. Rigid-receptor docking that incorporates CSA and a scoring function that modifies AutoDock scoring function is covered in Section 2. This initial method is also called LigDockCSA, and LigDockCSA now refers to the rdig-receptor mode of GalaxyDock. Development of the flexible-receptor mode of GalaxyDock is discussed in Section 3. Improvements of GalaxyDock by using beta complex and binding affinity function resulted in GalaxyDock2, as described in Section 4.

2. LigDockCSA: a rigid docking program

2.1. Overview of this section

As mentioned in introduction section, for PLD problem there has been many approaches that adopt an brand-new optimization techniques such as ant-colony algorithm to solve the problems of old techniques. For example, genetic algorithm might fall into local minimum easily rather than find global minimum. This phenomena causes difficulty of predicting correct binding mode when the energy landscape is sophisticated or number of torsion angles of ligand is very large (Chen *et al*., 2007). In this section, a program called LigDockCSA is introduced. It incorporates conformational space annealing as an optimization technique. It is known that CSA is a powerful global optimization technique (Lee *et al*., 1997) and it is applied to many computational biology field successfully (Lee *et al*., 2005a, Joo *et al*., 2009, Park and Seok, 2012). Scoring function of LigDockCSA is a modification of AutoDock3 scoring function (Morris *et al*., 1999) with adding torsion part of piecewise linear potential (PLP) (Gehlharr *et al*., 1995). When CSA searches ligand binding mode with AutoDock3 scoring function, it can find lower energy conformation when compared to Lamarkian genetic algorithm (LGA). However, deviations of the predicted conformation from native structure is more higher than that of LGA conformation. Adding PLP torsion term to AutoDock3 scoring function helps to find more accurate binding mode.

In the following subsections, details of LigDockCSA will be discussed.

2.2. Methods

2.2.1. Benchmark set and AutoDock calculation

The performance of the docking method developed in this work, LigDockCSA, is compared with two existing docking programs GOLD (Jones *et al*., 1997) and AutoDock (Morris *et al*., 1999) on the ASTEX diverse set (Hartshorn *et al*., 2007). The performance of GOLD on the set is reported in Hartshorn *et al*., and AutoDock calculations were carried out in this work. The ASTEX diverse set was developed to provide a suitable set for validation of docking methods. It was derived from high-resolution x-ray protein-ligand complex structures by focusing on the relevance to drug discovery and the quality of ligand structures. The set contains 85 diverse protein-ligand complexes that are all from different drug discovery or agrochemical targets. The associated ligands all meet drug-like criteria. The number of torsion angles of the ligands in the set ranges from 0 to 16.

For AutoDock and LigDockCSA calculations, the same protein and ligand input files and the same grid energies are used. Hydrogen atoms were added and Gasteiger-Hückel charges were assigned to proteins and ligands using Chimera (Petterson *et al*., 2004) after separating ligand and solvent molecules from the X-ray crystal complex structures in the ASTEX diverse set. A cubic grid box with the default

dimension of (22.5Å)3 was centered at the ligand atom closest to the geometrical center of the crystal ligand pose.

AutoDock results were obtained by executing the AutoDock3.0.5 program. For each complex, Lamarckian genetic algorithm (LGA) runs were performed thousand times to adjust the number of energy function calls to the same level as that for LigDockCSA which is on the order of 108.

2.2.2. Energy function for protein-ligand docking

First, it was tested that the performance of a docking protocol that combines CSA and the AutoDock energy. The AutoDock energy function, $E_{AutoDock}$, consists of four energy terms, Lennard-Jones interaction energy, directional hydrogen bond energy, screened Coulomb electrostatic potential energy, and solvation energy as follows:

$$E_{AutoDock} = w_{vdW}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + w_{hbond}\sum_{i,j}h(t_{ij})\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) \quad (2.1)$$
$$+ w_{elec}\sum_{i,j}\frac{q_i q_j}{\epsilon(r_{ij})r_{ij}} + w_{sol}\sum_{i,j}(S_i V_j + S_j V_i)\exp\left(-\frac{r_{ij}}{2\sigma^2}\right)$$

The torsional entropy term of AutoDock is omitted in Eq. (2.1) because it is independent of conformation. The same weight factors (w_{vdW}, w_{hbond}, w_{elec}, w_{sol}) used in AutoDock were adopted. Each of the above four energy terms is expressed as sum of interaction energy over all pairs of ligand atoms i and protein atoms j in addition to over all pairs of ligand atoms i and j that are separated by three or more chemical bonds. $\{A_{ij}, B_{ij}\}$ and $\{C_{ij}, D_{ij}\}$ are parameters for the Lennard-Jones and hydrogen bond energy, respectively. $h(t)$ is the weight factor for

hydrogen bond (H-bond) that depends on the angle t formed by the H-bond acceptor, H, and its donor (Goodford, 1985). q_i is partial charge, and $\varepsilon(r)$ is a distance-dependent dielectric constant. ($\{S_i, V_i\}$, σ) are solvation parameters.

In this study it was observed that many low-energy non-native structures are found by CSA, which disfavors native-like structures in terms of the AutoDock energy. This originates from the fact that the internal energy of the ligand molecule in AutoDock does not include proper torsional strains. Since the conformational flexibility of the ligand molecule is allowed in this study, torsional energy term from the PLP score was adopted (Gehlharr *et al.*, 1995)

$$E_{PLPtor} = \sum_{i=1}^{N_{tor}} F_k[1 + \cos(n_k\phi_k - \phi_{ko})] \tag{2.2}$$

For each rotatable bond k of ligand with torsion angle φk parameters are set to (F_k = 3.0, n_k = 3, ϕ_{ko} = π) for sp^3-sp^3 bond and (F_k = 1.5, n_k = 6, ϕ_{ko} = 0) for sp^2-sp^3 bond, as in the original PLP score. The final form of the energy function used in this study is expressed as a linear sum of the AutoDock energy function in Eq. (2.1) and the torsional energy term of the PLP score in Eq. (2.2) as follows:

$$E = E_{AutoDock} + wE_{PLPtor} \tag{2.3}$$

To determine the weight factor w in Eq. (2.3), a randomly divided half of the ASTEX diverse set was used as the training set and is listed in Supplementary table S1. Decoy conformations of the ligand for each complex in the training set were generated by perturbing the native conformation to varying magnitudes and selecting those with the

minimized energy lower than zero. After generating 2500 conformations and clustering them into 500 clusters by k-means clustering, the 500 cluster centers were chosen as decoys. Finally, w is set to 0.1 at which the energy-RMSD correlation of the decoy conformations for training set complexes becomes the maximum. This value of w is similar to those of other scaling factors used in Eq. (2.1) for the AutoDock energy which are also on the order of 0.1.

2.2.3. Application of conformational space annealing to protein-ligand docking

LigDockCSA determines the docking pose of a ligand by searching for the global minimum of the docking energy shown in Eq. (2.3) using conformational space annealing (CSA). CSA is an efficient global optimization technique that has been applied to various combinatorial optimization problems including protein folding and other biological problems (Lee *et al*., 2005a, Joo *et al*., 2009, Park and Seok, 2012). In CSA, a relatively small number (~100) of "bank" conformations is evolved by gradually reducing the effective size of the conformational space explored by each bank member. A distance measure in the conformational space is introduced as an annealing parameter for this purpose. During evolution of the bank, trial conformations are generated by crossovers and mutations as in genetic algorithm (Morris *et al*., 1999, Jones *et al*., 1995).

In this work, the conformational space is formed by 3 translational, 3 rotational, and N_{tor} torsional degrees of freedom for

ligand. The translational degrees of freedom are represented by the Cartesian coordinate of the ligand center atom, and the rotational degrees of freedom by the quaternion for the ligand orientation.

CSA searches for low energy conformations in the space of local energy minima, and local energy minimization is performed whenever a new conformation is generated. The simplex algorithm (Neider and Mead, 1965) is used, because it performs efficient minimization without requiring the gradient of the target function.

A CSA run starts with the generation of a "bank" of conformations, called the "first bank", and the bank conformations are evolved with iteration. We set the number of bank conformations to 50. In LigDockCSA, conformations for the first bank are generated by energy minimization of randomly produced conformations after filtering out those with initial energy $> 5 \times 10^5$ for which severe steric clashes occur. Only conformations with minimized energy lower than zero are included, as in the generation of the initial population in AutoDock.

At each iteration step of CSA, trial conformations are generated by crossovers and mutations, and the bank is updated considering both the current bank and the trial conformations. A trial conformation is generated by mixing a seed conformation selected from the current bank with a partner conformation randomly selected from the first bank or the current bank. The first bank helps to maintain conformational diversity during sampling, and the current bank helps to explore emerging low energy minima. For each of 30 seed conformations, 3 and 5 trial conformations are generated by mixing with the translation

and rotation of the conformations selected from the first bank and the current bank, respectively. Small random perturbations are introduced to the translation and rotation after mixing for more efficient sampling. Additional 2 and 5 trial conformations are generated for each seed by mixing with between 1 and $0.4N_{tor}$ torsion angles of the conformations selected from the first bank and the current bank, respectively.

After trial conformations are generated, the bank is updated considering structural diversity and energy of the current bank and the trial conformations. The essence of CSA is to focus on narrower conformational space of lower energy gradually as the iteration proceeds. For this purpose, a measure of distance between two conformations is required, and we adopt the distance measure employed by Lee *et al*. (2005b) except for requiring that the distance measures for translation, rotation, and torsion angles for the first bank have a ratio of $1:1:N_{tor}/6$ instead of 1:1:1. In CSA, the effective size of the conformational space represented by each bank member is controlled by the distance parameter D_{cut} at the stage of bank update. If a trial conformation within D_{cut} from a bank conformation has lower energy than the bank conformation, it replaces the bank conformation. If a trial conformation has distances greater than D_{cut} from all the current bank conformations and has lower energy than the highest energy bank conformation, it replaces the highest energy bank conformation. The parameter D_{cut} is gradually reduced as CSA iteration proceeds, and therefore conformational search focuses on narrower spaces of lower energy. When all bank members are used as seed, one round of CSA

terminates. Two rounds of CSA in LigDockCSA is executed. The second round of CSA starts with additional 50 randomly generated conformations in the first bank.

2.3. Results and discussion

2.3.1. Performance of CSA when combined with AutoDock scoring function

It has been demonstrated repeatedly that the CSA algorithm finds low energy solutions more efficiently than other popular global optimization methods such as Monte Carlo minimization, simulated annealing, etc. Here, the phenomena that the CSA can find lower energy minima than the LGA (Lamarckian genetic algorithm) implemented in AutoDock3.0.5 is shown when the same AutoDock energy function is employed. The docking procedure that adopts CSA and the AutoDock energy is referred to here as AutoDockCSA. As can be seen from the test results on the 85 complexes of the ASTEX diverse set presented in Table 2.1, AutoDockCSA tends to find lower energy conformations than AutoDock that uses LGA, especially for complexes with the number of rotatable bonds $N_{tor} \geq 5$. For example, when $N_{tor} \geq 12$, AutoDockCSA finds lower energy conformations than AutoDock in all 12 cases.

Table 2.1. The number of wins in finding lower energy conformations for AutoDock that employs LGA and AutoDockCSA that employs CSA when the identical AutoDock energy function is used

N_{tor}[a]	$N_{complex}$[b]	No. of wins			Lower E, higher RMSD[d]
		AutoDock (LGA)	AutoDockCSA (CSA)	Draw[c]	
0~4	24	8	10	6	12
5~7	26	4	21	1	17
8~11	23	8	15	0	16
12~16	12	0	12	0	11
Total	85	20	58	7	56

a. The number of rotatable bonds of the ligand

b. The number of complexes in the ASTEX diverse set with Ntor in the specified range

c. The number of cases in which AutoDock and AutoDockCSA give the same energy when the energy value is truncated at two digits below the decimal point, as in the AutoDock output

d. The number of cases in which either LGA or CSA finds a lower energy but higher RMSD conformation

Although CSA can find lower energy conformations than LGA on average, the docking results judged by RMSD from the native pose are not improved by AutoDockCSA. This should be viewed in contrast to the recent studies (Janson *et al*., 2008, Chen *et al*., 2007) that improved optimization of the AutoDock energy led to enhanced docking results. The success rate defined by the percentage of the cases in which RMSD of the lowest energy conformations is less than 2 Å is summarized in Table 2.2. The overall success rate for the ASTEX diverse set is 81.7% for AutoDock and 68.2% for AutoDockCSA. The poor performance of AutoDockCSA is due to the fact that lower energy conformations often have higher RMSD with the AutoDock energy, as shown in Table 2.2. It is also notable that the proportion of such cases increases with the increasing number of ligand torsions. This implies that the AutoDock energy has a problem that becomes apparent only when a more rigorous global optimization method is used, in contrast to the previous reports that better optimization of the AutoDock energy improves the overall prediction accuracy (Janson *et al*., 2008, Chen *et al*., 2007). To overcome this problem, the AutoDock energy function is improved in order to take full advantage of the more efficient global optimization method.

Table 2.2. Success rate and average RMSD of AutoDock and AutoDockCSA for the ASTEX diverse set

N_{tor}	$N_{complex}$	Success rate[a] (AverageRMSD[b])	
		AutoDock (LGA)	AutoDockCSA (CSA)
0~4	24	75.0% (1.80 Å)	66.7% (2.12 Å)
5~7	26	88.5% (1.24 Å)	80.8% (1.84 Å)
8~11	23	82.6% (1.56 Å)	60.7% (2.34 Å)
12~16	12	75.0% (2.08 Å)	58.3% (3.09 Å)
Total	85	81.7% (1.60 Å)	68.2% (2.23 Å)

a. Percentage of the cases in which RMSD of docked pose is less than 2 Å

b. Average of RMSD of the lowest energy conformations over the complexes in the set

2.3.2. Energy function for LigDockCSA

The energy function adopted here, $E = E_{AutoDock} + wE_{PLPtor}$, has a free weight parameter w. This parameter was determined so that the maximum correlation between the energy and RMSD for decoys in the training set is achieved, as explained in Methods. The average Pearson's correlation coefficient over the training set complexes is 0.402 when the torsional energy is not included ($w = 0$) and reaches the maximum value of 0.462 when w is 0.1, as can be seen from Figure 2.1. This rather small improvement may be related to the small value of the weight factor. The correlation coefficient for each of the complexes in the training set is presented in Table 2.3.

To test the sensitivity of the scaling factor w, on the training set, the training set was randomly divided into three subsets of 14 complexes, and average correlation coefficient for each subset was examined. The correlation maximum appears at $w = 0.1$ for two subsets (the next highest correlations at 0.0 and 0.2) and at $w = 0.6$ for the other subset (the next highest correlation at 0.0). It can be therefore expected that the current parameter of $w = 0.1$ would not be very sensitive to the choice of the training set.

Figure 2.1. The average of the energy-RMSD correlation coefficients for the training set decoys reaches maximum at $w = 0.1$.

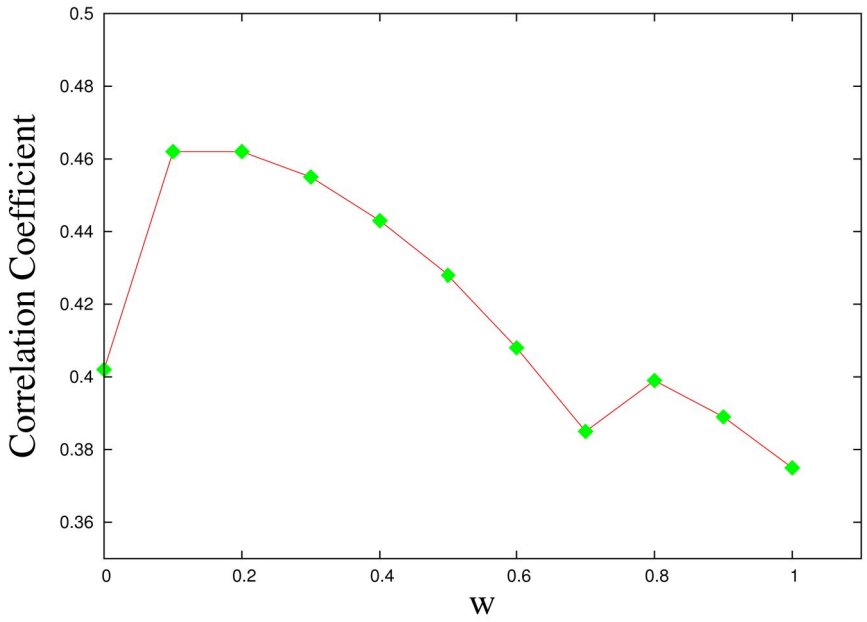


Table 2.3. Pearson's correlation coefficients between Energy and RMSD for the training set

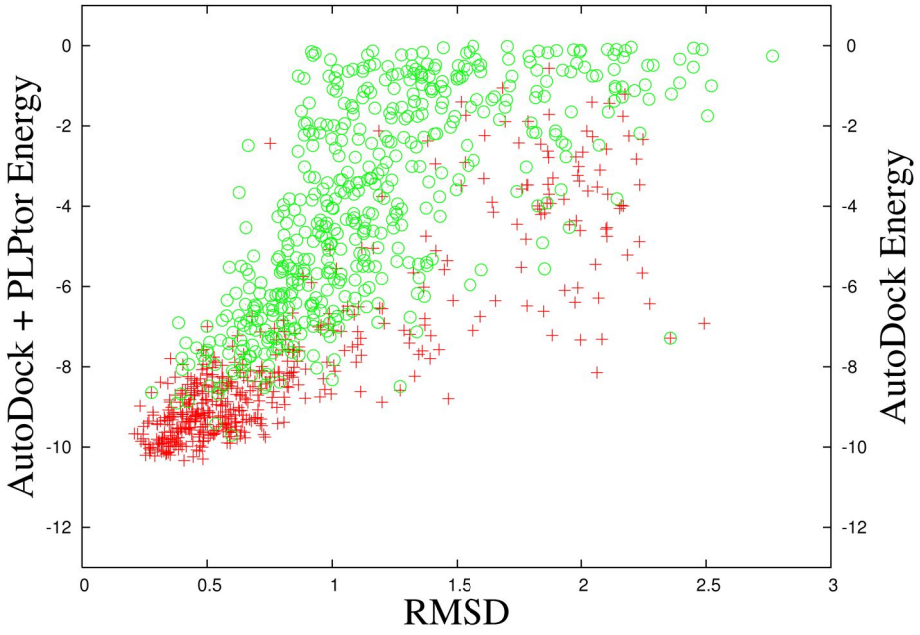
PDB	N_{tor}	ADE[a]	LDE[b]	PDB	N_{tor}	ADE[a]	LDE[b]
1q41	1	0.187	0.744	1lpz	7	0.409	0.579
1hnn	1	0.394	0.406	1u1c	7	0.228	0.291
1p2y	1	0.448	0.385	1xoq	7	0.540	0.755
1uou	2	0.479	0.580	2bm2	7	0.679	0.681
1gpk	3	0.618	0.632	1m2z	8	0.667	0.520
1hww	3	0.452	0.546	1sj0	8	0.208	0.244
1sg0	3	0.442	0.725	1xm6	8	0.505	0.204
1ia1	4	0.457	0.506	1z95	8	0.503	0.488
1l2s	4	0.354	0.287	1n46	9	0.279	0.274
1oyt	4	0.625	0.814	1s19	9	0.517	0.572
1r9o	4	0.176	0.394	2bsm	9	0.514	0.498
1yvf	4	0.478	0.586	1meh	10	0.160	0.252
1ke5	5	0.524	0.535	1sq5	10	0.285	0.242
1of1	5	0.158	0.131	2br1	10	0.360	0.349
1p62	5	0.325	0.328	1uml	11	0.197	0.374
1v4s	5	0.342	0.213	1hvy	12	0.602	0.558
1x8x	5	0.131	0.336	1mzc	12	0.195	0.181
1hp0	6	0.300	0.516	1r58	12	0.082	0.149
1opk	6	0.516	0.155	1unl	12	0.495	0.744
1q1g	6	0.376	0.672	1ygc	13	0.471	0.455
1t9b	6	0.559	0.674	1gkc	16	0.640	0.800

a. ADE: AutoDock energy

b. LDE: LigDock energy, $E_{AutoDock} + 0.1E_{PLPtor}$

An interesting example of the improved cases by the addition of the PLP torsion energy is shown in Figure 2.2 for the complex 1Q1G. The energy-RMSD correlation coefficient for the decoys for this complex is 0.376 with the AutoDock energy and 0.672 with the additional torsion energy. It should be noted that the energy and RMSD for decoys are evaluated after minimization of the new energy function. The lowest energy conformation moves slightly towards the native from RMSD of 0.617 Å with the AutoDock energy to 0.372 Å with the new energy function. A large portion of decoy conformations also move towards the native structure, as can be seen from Figure 2.2. This fact implies that the relief of torsional strain in the decoys by local minimization of the new energy function drives structural changes towards the native structure. Examination of the ligand torsion angles of the decoys confirmed that the torsion angles indeed converge to the native angles upon energy minimization (data not shown).

Figure 2.2. Modification of the energy landscape of 1Q1G is illustrated after the addition of the PLP torsion energy. The AutoDock energy of the decoy is shown in green circles, and the AutoDock energy combined with the PLP torsion energy is shown in red crosses. Overall shift towards smaller value of RMSD after adding the PLP torsion energy can be observed as torsional strains are relieved by local energy minimization. The unit of *x* axis is kcal/mol.



2.3.3. Performance of LigDockCSA

The LigDockCSA program predicts the docking pose of a ligand by finding low energy minima of the energy function in Eq. (2.3) using CSA. RMSDs of the predicted poses by LigDockCSA for individual complexes are reported in Supplementary table S2. The performance of LigDockCSA was tested on the ASTEX diverse set. Because one half of the complexes (42 out of 85 complexes) were used to determine the energy parameter *w*, we first examined the performance of LigDockCSA on the training set (See Table 2.4) and that for the other half of the ASTEX set, which we call test set (See Table 2.5) separately. We also compare the results with the performance of GOLD and AutoDock in the tables. From the results presented in Tables 2.4 and 2.5, one can conclude that the energy parameter w is not over-trained since the performance of LigDockCSA on the training set is not particularly superior to that on the test set.

For comparison with GOLD, the results of the standard protocol of GOLD out of the six methods reported in Hartshorn *et al.* are presented since the standard protocol is the most similar to the current AutoDock and LigDockCSA calculations in the size of the binding site and in the preparation of protein and ligand molecules. However, the results of GOLD should be taken with a notion that the number of function evaluations in GOLD calculations is not known.

Table 2.4. The success rates and the average RMSD values of GOLD, AutoDock, AutoDockCSA, and LigDockCSA are shown for the training set of 42 complexes.

N_{tor}	$N_{complex}$	Success rate (Average RMSD)			
		GOLD[a]	AutoDock	AutoDock CSA	LigDock CSA
0~4	12	66.7%	83.3% (1.47 Å)	58.3% (2.52 Å)	91.7% (1.24 Å)
5~7	13	95.0%	92.3% (1.11 Å)	84.6% (1.39 Å)	92.3% (1.10 Å)
8~11	11	73.2%	81.8% (1.42 Å)	54.5% (2.30 Å)	72.7% (1.29 Å)
12~16	6	62.5%	66.7% (2.50 Å)	33.3% (3.93 Å)	66.7% (1.81 Å)
Total	42	76.5%	83.3% (1.49 Å)	61.9% (2.31 Å)	83.3% (1.29 Å)

a. The success rate is reported in Hartshorn *et al*., but the average RMSD is not available.

Table 2.5. The success rates and the average RMSD values of GOLD, AutoDock, AutoDockCSA, and LigDockCSA are shown for the test set of 43 complexes.

N_{tor}	$N_{complex}$	Success rate (Average RMSD)			
		GOLD[a]	AutoDock	AutoDockCSA	LigDockCSA
0~4	12	76.7%	66.7% (2.13 Å)	75.0% (1.72 Å)	81.8% (1.12 Å)
5~7	13	83.5%	84.6% (1.36 Å)	72.7% (2.28 Å)	84.6% (1.78 Å)
8~11	12	85.0%	83.3% (1.68 Å)	66.7% (2.38 Å)	91.7% (1.11 Å)
12~16	6	100.0%	83.3% (1.65 Å)	63.3% (2.25 Å)	83.3% (1.59 Å)
Total	43	84.3%	81.0% (1.56 Å)	69.8% (2.15 Å)	85.7% (1.42 Å)

a. The success rate is reported in Hartshorn *et al*., but the average RMSD is not available.

As presented in Table 2.5, the success rate for the test set with the RMSD threshold of 2 Å improves from 81.0% for AutoDock to 85.7% for LigDockCSA. The average RMSD also improves from 1.56 Å for AutoDock to 1.42 Å for LigDockCSA. For a reference, the overall success rate for the 85 complexes of the full ASTEX set is 80.5%, 81.7%, and 84.7% for GOLD, AutoDock, and LigDockCSA, respectively, and the average RMSD is 1.60 Å and 1.35 Å for AutoDock and LigDockCSA, respectively (See Table 2.6). The success rate of LigDockCSA is the same as that of AutoDock for the number of rotatable bonds of 5~7 and 12~16, but the average RMSD improves for these complexes. LigDockCSA tends to fail (i.e., RMSD > 2 Å) on the same complexes as AutoDock does, 10 out of 13 failures being shared with those of AutoDock. This is probably due to the fact that the LigDockCSA energy function is not so much different from the AutoDock energy function except that it contains an additional torsional energy term and the overall energy landscapes are similar to each other. It is still noticeable that the average RMSD improves with LigDockCSA. The percentage of the predictions within 1 Å RMSD for the full ASTEX set improved from 47.0% for AutoDock to 56.5% for LigDockCSA. This implies that the LigDockCSA energy landscape is improved in fine structural details even when the overall energy landscape is similar to that of AutoDock. Considering that CSA was not successful with the original AutoDock energy function, such refinement in the energy landscape is essential to take full advantage of the efficient global optimization method.

Table 2.6. Success rate and average RMSD of GOLD, AutoDock, and LigDockCSA for the full ASTEX diverse set

N_{tor}	$N_{complex}$	Success rate (Average RMSD)		
		GOLD[a]	AutoDock	LigDockCSA
0~4	24	71.7%	75.0% (1.80 Å)	87.5% (1.20 Å)
5~7	26	89.2%	88.5% (1.24 Å)	88.5% (1.44 Å)
8~11	23	79.3%	82.6% (1.56 Å)	82.6% (1.21 Å)
12~16	12	81.2%	75.0% (2.08 Å)	75.0% (1.68 Å)
Total	85	80.5%	81.7% (1.60 Å)	84.7% (1.34 Å)

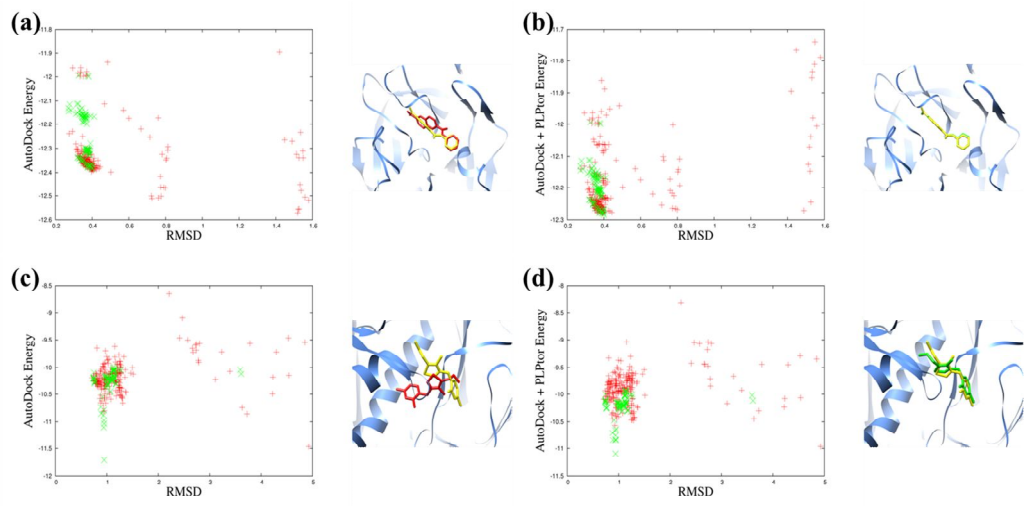
a. The success rate reported in Hartshorn *et al*.. Average RMSD is not reported in the reference.

Two examples, 1OWE and 1IG3, in which LigDockCSA produces improved predictions are illustrated in Figure 2.3. In the case of 1OWE, the improvement is due to the improved energy landscape by addition of the torsional energy term. For 1IG3, the energy landscape does not change much by the torsional energy, but the conformational search by CSA that generates lower energy conformations contributes to the improvement.

We further examined the effect of clustering by selecting the lowest energy conformation of the largest cluster after clustering the final conformations using NMRCLUST (Kelley *et al*., 1996). The docking performance is improved by clustering: the overall success rate improves to 88.2% and 89.4%, and the average RMSD improves to 1.19 Å and 1.05 Å for AutoDock and LigDockCSA, respectively. These results imply that a bound state corresponding to a broader energy basin tends to be favored, which is in accordance with recent reports that illustrated the importance of incorporating conformational entropy to generate more precise prediction of binding poses (Lee and Seok, 2009, Lu and Wong, 2005, Ruvinsky, 2007).

Despite the success in the pose prediction, the new method improves the prediction accuracy of binding affinity very little compared to AutoDock because the magnitude of the added torsion term is very small (data not shown).

Figure 2.3. Energy landscapes and binding poses obtained by AutoDock and LigDockCSA for two complexes 1OWE (N_{tor} = 3) and 1IG3 (N_{tor} = 8). Energy is plotted with respect to RMSD for the 100 CSA final bank conformations in green 'x' and the 250 conformations from the AutoDock run that contains the lowest AutoDock energy conformation in red '+'. The ligand poses predicted by AutoDock are shown in red and those by LigDockCSA in green. The native ligand poses are colored in yellow. (a) AutoDock energy and the binding pose for 1OWE (RMSD = 1.5 Å) predicted by AutoDock, (b) the new energy with the additional PLP torsion energy and the binding pose for 1OWE (RMSD = 0.4 Å) predicted by LigDockCSA, (c) AutoDock energy and the binding pose for 1IG3 (RMSD = 4.9 Å) predicted by AutoDock, and (d) the new energy with the additional PLP torsion energy and the binding pose for 1IG3 (RMSD = 0.9 Å) predicted by LigDockCSA.



2.4. Conclusion of this section

In this section, the powerful global optimization technique of CSA with the popular docking energy function from AutoDock have combined. It has been shown that the straightforward application of the efficient CSA search method to the current AutoDock energy identifies a loophole in the scoring function, and the scoring function is modified to include a proper torsional energy term. Consequently more precise prediction of binding poses than AutoDock and GOLD became possible when the CSA search method was applied to the refined AutoDock energy function. From this experience, one can expect that further improvements in the scoring function may facilitate additional improvements in docking performance when more rigorous conformational search methods are combined.

The importance of efficient search methods coupled with reasonably accurate energy functions becomes even more apparent when the conformational search space grows, for example, to include the flexibility of the protein binding site, which will be provided in next section. The feedback between the search method and the proper modification of the energy function is critical for such applications. The current approach provides a proof-of-principle study for such an extension.

3. GalaxyDock: a flexible receptor docking

3.1. Overview of this section

One of important issues of protein-ligand docking is incorporating receptor flexibility. As described in section 1.2, there has been many approaches to develop flexible docking algorithm. Some of them use conformational ensemble, others dock ligand to enlarged binding pocket, and the others employ simultaneous approach.

In this section, a docking program called GalaxyDock, which belongs to the third class of flexible docking methods is introduced. This method is an extension of LigDockCSA, a previous version that treats the protein as rigid and the ligand as flexible, and which outperforms AutoDock3 (Morris *et al.*, 1999) and GOLD (Jones *et al.*, 1995, 1997) when tested on the ASTEX diverse set as shown in Section 2. We attribute the success of LigDockCSA to both the effectiveness of the CSA to a more refined docking energy. When compared to AutoDock3, lower energy conformations could be identified by the powerful sampling with CSA when the same AutoDock energy was used. However, those conformations were more distant from the experimental structures on average. This problem was corrected by adding the ligand torsional energy of Piecewise Linear Potential (PLP). Here, GalaxyDock extends the CSA global optimization employed in LigDockCSA for efficient sampling of protein side-chain flexibilities as

well as ligand flexibilities. GalaxyDock employs the same energy for protein-ligand and intra-ligand interaction energy as LigDockCSA. Intra-protein interaction energy is incorporated to consider protein side-chain conformational change by combining the ROTA score (Hartmann *et al.*, 2007) and van der Waals energy after training on the ASTEX diverse set. GalaxyDock shows improved performance over LigDockCSA and other flexible docking programs when tested on 3 sets involving protein side-chain conformational changes in the binding pocket, Human Immunodeficiency Virus protease (HIV-PR), Liver X Receptor beta (LXRβ), and cAMP-dependent protein kinase (cAPK). When tested on a more diverse set of 16 proteins, the performance of GalaxyDock is comparable to the SCARE method (Bottegoni *et al.*, 2008). This result implies that simultaneous sampling of protein and ligand conformations by powerful global optimization of a carefully designed docking energy may be a promising avenue for future development of more advanced protein-ligand docking methods.

3.2. Methods

3.2.1. GalaxyDock energy function for flexible protein-ligand docking

The GalaxyDock energy function is expressed as follows:

$$E = E_{protein-ligand} + E_{intra-ligand} + E_{intra-protein} \quad (3.1)$$

where the first and the second terms, protein-ligand and intra-ligand interaction energy, are taken from LigDockCSA, which is a linear combination of AutoDock3 energy function and the ligand torsion energy of PLP scoring function. The third term of Eq. (3.1) represents intra-protein energy. When the energy used to generate the receptor energy grid in LigDockCSA is used for the intra-protein energy, a rather poor performance was obtained in the initial tests, often failing in reproducing protein-ligand hydrogen bonds (unpublished data).

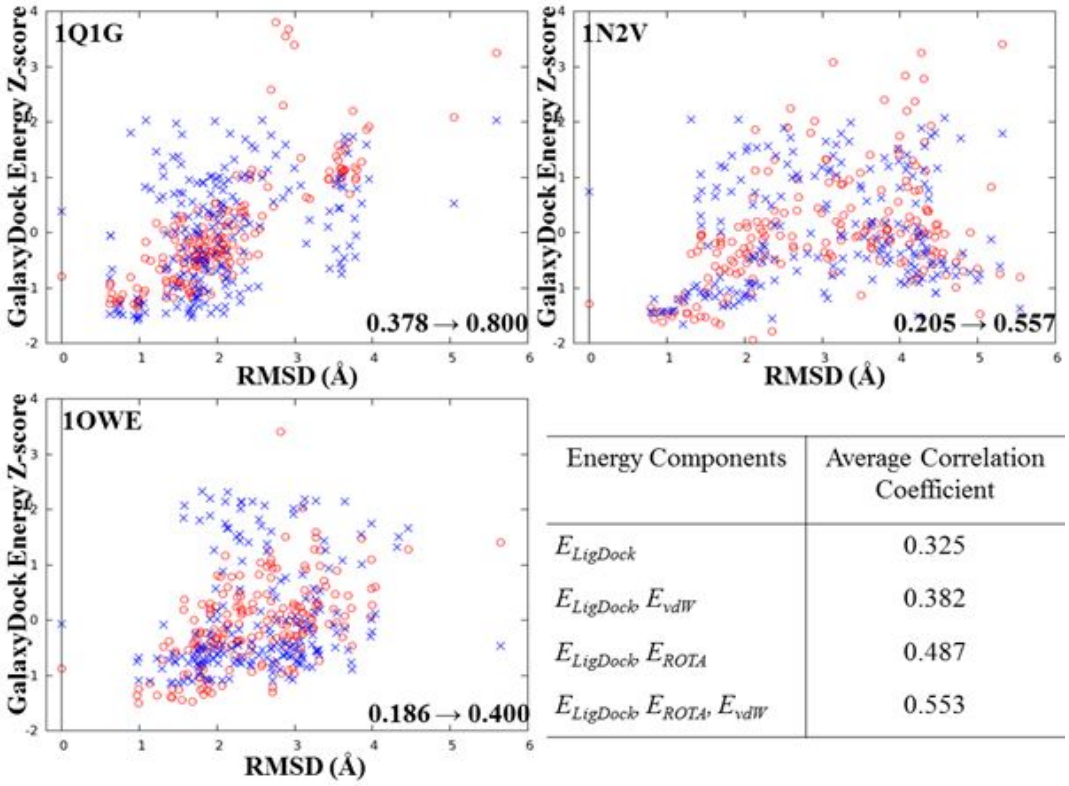
To solve the problem of describing intra-protein energy more accurately in flexible docking, the ROTA energy function (Hartmann *et al.*, 2007) is incorporated. ROTA is a statistical potential that was reported to show reliable performance in predicting side-chain orientation and in rescoring protein-ligand binding poses (Hartmann *et al.*, 2009). We noted that the use of ROTA alone on occasion fails to filter out steric clashes between side-chains, and consequently the van der Waals term was added to prevent bumps, as follows:

$$E_{intra-protein} = w_{ROTA}E_{ROTA} + w_{vdW}E_{vdW,intra-protein} \quad (3.2)$$

To reduce the calculation time of this term, an energy grid in the space of side-chain torsion angles is generated in advance and the energy is evaluated by linear interpolation.

The weight parameters in Eq. (3.2) were determined to be w_{ROTA} = 0.06 and w_{vdW} = 0.05 by training on a decoy set for 85 complexes of the ASTEX diverse set to obtain maximum average Pearson's correlation coefficient between the energy and RMSD from the native structure (See Figure 3.1 for more details).

Figure 3.1. Energy-RMSD plots of 3 complexes from the ASTEX diverse set. Red "o" and blue "x" represent the decoy Z-score of the GalaxyDock energy and that of the LigDockCSA energy, respectively, with respect to RMSD of flexible residues and ligand from the crystal structure. GalaxyDock shows higher energy-RMSD correlation and lower energy for the native pose (RMSD = 0) than LigDockCSA. Change in the Pearson's correlation coefficient is shown on the lower right corner in each panel. Shown on the lower right corner are the average energy-RMSD correlation coefficients for different combinations of energy components.



Energy Components	Average Correlation Coefficient
$E_{LigDock}$	0.325
$E_{LigDock}$, E_{vdW}	0.382
$E_{LigDock}$, E_{ROTA}	0.487
$E_{LigDock}$, E_{ROTA}, E_{vdW}	0.553

Decoy structures were generated by randomly perturbing ligand conformations from the crystal structures and building side-chain conformations from Dunbrack's backbone dependent rotamer library (Dunbrack and Cohen, 1997) in proportion to the rotamer probabilities for the 3 nearest residues from the ligand geometrical center. According to Najmanovich *et al.* and Gaudreault *et al.*, about 85% of protein-ligand complexes in PDB show conformational changes in 3 residues or less. If the energy of a generated conformation was lower than 0 kcal/mol by LigDock score, which corresponds to no net interaction, the conformation was accepted as a decoy. This procedure was iterated until the number of accepted conformations reached 200 for each complex.

3.2.2. GalaxyDock sampling that incorporates side-chain flexibility

GalaxyDock is an extension of LigDockCSA that treats the receptor as rigid, and the conformational sampling is based on the same CSA global optimization. CSA has also been successfully applied to many other global optimization problems. In the CSA procedure, a fixed number of docking conformations (30 in this work) referred to "bank" is generated and evolved until convergence. A bank member represents the lowest energy conformation in the local conformational space of radius D_{cut} around it. The bank is updated after generating trial conformations by crossover and mutation at each iteration step. D_{cut} is gradually decreased during iteration, leading to annealing in the

conformational space. More details of the method are described below.

GalaxyDock starts with the generation of the initial bank. Conformations are generated by assigning a random value in a preset range to each component of the ligand coordinate (Cartesian coordinates for the ligand center atom, quaternion representing ligand rotation, and free torsion angles of ligand), and side-chain conformations of the flexible residues are taken from the Dunbrack backbone dependent rotamer library (Dunbrack *et al.*, 1997) in proportion to the rotamer probability. Only those conformations with initial energy < 500,000 kcal/mol, which corresponds to the energy of five strong clashes, are subject to simplex minimization, and only those with minimized energy < 0 kcal/mol are accepted. This procedure is repeated until the number of accepted conformations reaches 30 or the number of trials is 5,000.

A measure for the distance between two conformations is necessary to define D_{cut}, and GalaxyDock takes a linear combination of D_{trans}, the distance between ligand center atoms, D_{rot}, the angle between two vectors representing the Euler angle of ligand rotation, D_{tor}, the Hamming distance of ligand torsion angles, and D_{χ}, the Hamming distance of side-chain torsion angles. The relative weights between these components are determined such that the ratio of the average distances of the bank members is $D_{trans}:D_{rot}:D_{tor}:D_{\chi} = 1:1:N_{tor}/6:N_{\chi}/6$ in the initial bank, where N_{tor} is a number of free torsions in ligand, and N_{χ} is the number of flexible protein side-chain torsion angles.

At each CSA iteration step, trial conformations are generated by crossover or mutation of current bank members, as described in

detail for LigDockCSA. In GalaxyDock, protein conformations, as well as ligand conformations, are also mixed. The bank is updated by considering the energy and distance of the trial conformations with the previous bank members as follows: a trial conformation with lower energy than a previous bank member within D_{cut} replaces the previous member, and a trial conformation with distance > D_{cut} from all previous members replaces the highest-energy bank member. In this way, the size of the bank is kept constant, and the conformational search is focused on narrower areas with lower energy in the conformational space by gradually reducing D_{cut} with iteration. The iteration terminates when no new bank members are found, reaching convergence.

3.2.3. Cross-docking benchmark test

The performance of GalaxyDock was first tested by N x N cross-docking experiments, where N is the number of complexes, on the following 3 proteins: Human Immunodeficiency Virus (HIV) protease (N = 20), Liver X Receptor beta (LXRβ) (N = 3), and cAMP-dependent Protein Kinase (cAPK) (N = 4). In cross-docking, a ligand (say, ligand A) is docked to a structure (say, protein structure B) of the same protein bound to a different ligand (ligand B), and the docked pose is compared with the experimental bound structure (protein structure A + ligand A). Additional cross docking tests were performed on a more diverse set of 30 complexes involving 16 proteins that were collected by Bottegoni *et al.*.

Protein and ligand input files were prepared using Sybyl 8.1.

Hydrogen atoms were added, and Gasteiger-Hückel charges were assigned to each protein and ligand. A cubic grid box of (22.5 Å)^3 centered at the ligand geometrical center of the crystal binding mode was generated for the first three test sets. Grid box size for the diverse set is explained below. Test set structures and flexible residues for each test case were selected as follows.

3.2.3.1. HIV protease

HIV protease is required for the proper assembly and maturation of infectious virons (Brik *et al.*, 2003). If HIV protease is rendered ineffective, the HIV virus becomes uninfectious. Österberg *et al*. performed flexible docking using an ensemble of grids generated from 21 crystal structures of HIV protease bound to different ligands. We made a cross-docking test on 20 complex structures, excluding 1HVR. The ligand of 1HVR occupies the position of the water molecule present in other 20 complexes (Zhao and Sanner, 2007). Successful cross docking of this ligand requires proper treatment of the binding site water, but this is beyond the current scope. The PDB IDs of the 20 complexes are 1HBV, 1HEF, 1HEG, 1HIH, 1HIV, 1HPS, 1HTE, 1HTF, 1HTG, 1HVI, 1HVJ, 1HVK, 1HVL, 1HVS, 1SBG, 4HVP, 4PHV, 5HVP, 8HVP, and 9HVP. Flexible torsion angles of ligands are selected following Österberg *et al*. First, 2 residues, ARG8 of the 2 chains of HIV protease, and next, 4 residues, ARG8 and ILE50 of the 2 chains, were treated as flexible in 2 cross-docking tests since steric clashes caused by swapping ligands were the largest on

ARG8 and then on ILE50 in the previous study of Morris *et al.* (2009).

3.2.3.2. LXRβ

LXRβ, a member of the nuclear receptor superfamily, regulates genes involved in cholesterol and lipid metabolism (Färnegårdh *et al.*, 2003). Out of 6 crystal structures for LXRβ in PDB, 3L0E was eliminated because it is bound to a potent antagonist while the others are bound to agonists. 1UPV and 1PQ9 were excluded because they contain the same ligand as 1PQC, which has a better resolved structure. The 3 remaining complexes in the LXRβ test set are 1P8D, 1PQ6, and 1PQC. A PHE triad (PHE271, PHE329, and PHE340) in the binding pocket is treated as flexible since these residues show the largest conformational changes when the crystal structures are superimposed, consistent with previous crystal structure and molecular dynamics (Beautrait *et al.*, 2008) studies.

3.2.3.3. cAPK

cAPK, also known as protein kinase A (PKA), belongs to a family of enzymes whose activity depends on the cellular level of cyclic AMP. It performs multiple functions related to the regulation of sugar and lipid metabolism (Engh *et al.*, 1996). Huang and Zou attempted ensemble docking with a modified version of UCSFDock. Out of the 7 crystal structures in the Huang set, the following 3 complexes that have similar ligands were excluded since even rigid

docking can give successful cross-docking results for them: 1FMO (similar to 1BKX) and 1YDR and 1YDS (similar to 1YDT). The remaining 4 PDB structures, 1BKX, 1BX6, 1STC, and 1YDT, form the cAPK set. Two residues, PHE54 and PHE327, located at 2 opposite sides of the binding pocket show severe clashes with ligands when different ligand-bound crystal structures are superimposed. This selection for the flexible side-chains is also identical to that in the FLIPDock study of Zhao and Sanner.

3.2.3.4. Diverse set

An additional test set of 16 different proteins presented in the SCARE (Bottegoni *et al.*, 2008) paper was employed. This set is composed of 14 proteins bound to two different ligands and 2 proteins (aldose reductase and anti-steroid fab) unbound and bound to a ligand. Therefore, 30 cross-docking experiments were performed. All of the highly displaced or clashing residues reported in Table 2 of Bottegoni *et al.* (2008) were treated flexible. (See Table 3.1) Two types of binding pockets were tested to compare with the SCARE method. First, grid boxes were generated such that they include all the residues within 5 Å from the ligand atoms in the experimental structure, following the SCARE paper. Second, grid boxes were generated such that they include all the residues predicted to contact ligands by an in-house method called GalaxySite (http://galaxy.seoklab.org/site). GalaxySite is a ligand-binding site prediction method that performs docking simulations of predicted ligands selected based on similarity. The docking

simulations optimize a hybrid energy of AutoDock and template-derived restraints. It ranked the fourth place in the binding site prediction category in the 9th Critical Assessment of Techniques for Protein Structure Prediction when applied to binding site prediction of protein model structures (Schmidt *et al.*, 2011). For a fair comparison on the current test set, GalaxySite was run deleting the proteins in the diverse set from the list of proteins for ligand prediction. The SCARE method was used with ICMPocketFinder.

3.3. Results and discussion

Performance of GalaxyDock on the four test sets are compared with those of rigid receptor docking methods and other flexible docking methods in Table 3.2. Success rate and average RMSD are reported in the table. Success rate is defined as percentage of the cases in which ligand RMSD of the top scoring ligand pose is within 2 Å from the crystal complex structure. The average RMSD is the average of RMSDs of the top scoring ligand poses from those of the experimental structures. Results excluding self-docking cases are also shown in parenthesis if self-docking cases are included in the success rate or average RMSD to compare with other methods. Discussions on the results for each test are provided below.

Table 3.1. Selected residues of the diverse set

Protein	Receptor	Ligand	Flexible residues
Aldose Reductase	2ACR	2FZB	F122, L300
Anti-Steroid Fab	1DBA	1DBB	W100
CDK2	1AQ1	1DM2	H84
	1DM2	1AQ1	I10, H84, Q131, L148
COX-2	1CX2	3PGH	R120
	3PGH	1CX2	F518
Estrogen Receptor	1ERR	3ERT	E419, M421, L525
	3ERT	1ERR	M421, L525, L539
Factor Xa	1KSN	1XKA	Y99
	1XKA	1KSN	Q192
GSK-3 β	1Q4L	1UV5	R141
	1UV5	1Q4L	S66, R141, Q185
Hiv1 RT	1C1C	1RTH	K102, W229, P236
	1RTH	1C1C	K102, W229, P236
JNK3	1PMN	1PMV	I124, M146
	1PMV	1PMN	I124, M146
LXR β	1P8D	1PQ6	L330, F340, I353
	1PQ6	1P8D	R319, I353
Neuroaminidase	1A4Q	1NSC	E274
	1NSC	1A4Q	E274
P38 Kinase	1BMK	1DI9	M109
	1DI9	1BMK	I84, M109
PKA	1STC	1YDS	T183, F327
	1YDS	1STC	F54, F327
PPARγ	1FM9	2PRG	F282, F363
	2PRG	1FM9	E259, F282, Q286, F363, L453
TK	1KI4	1KIM	Y101, Y132, Y172
	1KIM	1KI4	Y101, Y132
Trypsin	1PPC	1PPH	Q192
	1PPH	1PPC	Q192, Y228

Table 3.2. Comparison of cross-docking test results in terms of success rate and average RMSD of the top scoring poses on the HIV protease set, LXRβ set, cAPK set, and diverse set for GalaxyDock and other methods

Set	Program	Method	Success rate[a] (%)	Average RMSD[b] (Å)
HIV protease set	GalaxyDock	2 flexible res[e]	98.8 (98.9)	1.03 (1.03)
		4 flexible res[f]	98.0 (98.2)	1.02 (1.03)
	LigDockCSA	rigid docking	75.0 (73.7)	2.76 (2.89)
	FLIPDock[c]	2 flexible res[e]	96.3	1.09
		8 flexible res[g]	93.5	1.12
	AutoDock3[d]	ensemble grid	87.0	-
		rigid docking	72.0	-
LXRβ set	GalaxyDock	3 flexible res[i]	88.9 (83.3)	1.63 (1.91)
	LigDockCSA	rigid docking	22.2 (0.0)	3.54 (4.40)
	RossetaLigand[h]	flexible sc[j]	55.6	1.91
		flexible sc+bb[k]	44.4	1.84
	AutoDock4[h]	rigid docking	22.2	3.14
cAPK set	GalaxyDock	2 flexible res[l]	62.5 (58.3)	2.07 (2.27)
	LigDockCSA	rigid docking	43.8 (25.0)	3.89 (4.83)

Diverse set	GalaxyDock	known pocket[n]	86.7	1.70
		predicted pocket[o]	80.0	2.05
	LigDockCSA	known pocket	43.3	3.17
		predicted pocket	46.7	2.76
	SCARE[m]	known pocket[n]	90.0	-
		predicted pocket[p]	80.0	-

a. Percentage of the cases in which RMSD of top scoring ligand is < 2 Å from the crystal complex structure and that excluding self-docking cases in parenthesis. Diverse set does not contain self-docking cases.

b. Average ligand RMSD for top scoring ligands over all cases and that excluding self-docking cases in parenthesis

c. Data taken from Tables 1 and 2 of Zhao and Sanner

d. Data taken from Österberg *et al.*

e. ARG8 of the two chains are treated flexible.

f. ARG8 and ILE 50 of the two chains are treated flexible.

g. ARG8, ASP29, ASP30, and VAL82 of the two chains are treated flexible.

h. Data taken from Table 1 of Davis and Baker

i. PHE triad is selected to be flexible.

j. All side-chains in the binding pocket are treated flexible.

k. Limited backbone flexibility is allowed during energy minimization after side-chain optimization.

l. PHE54 and PHE327 are treated flexible.

m. Data taken from Tables 6 and 8 of Bottegoni *et al.* (2008)

n. Pocket is composed of all residues within 5 Å from any ligand atom

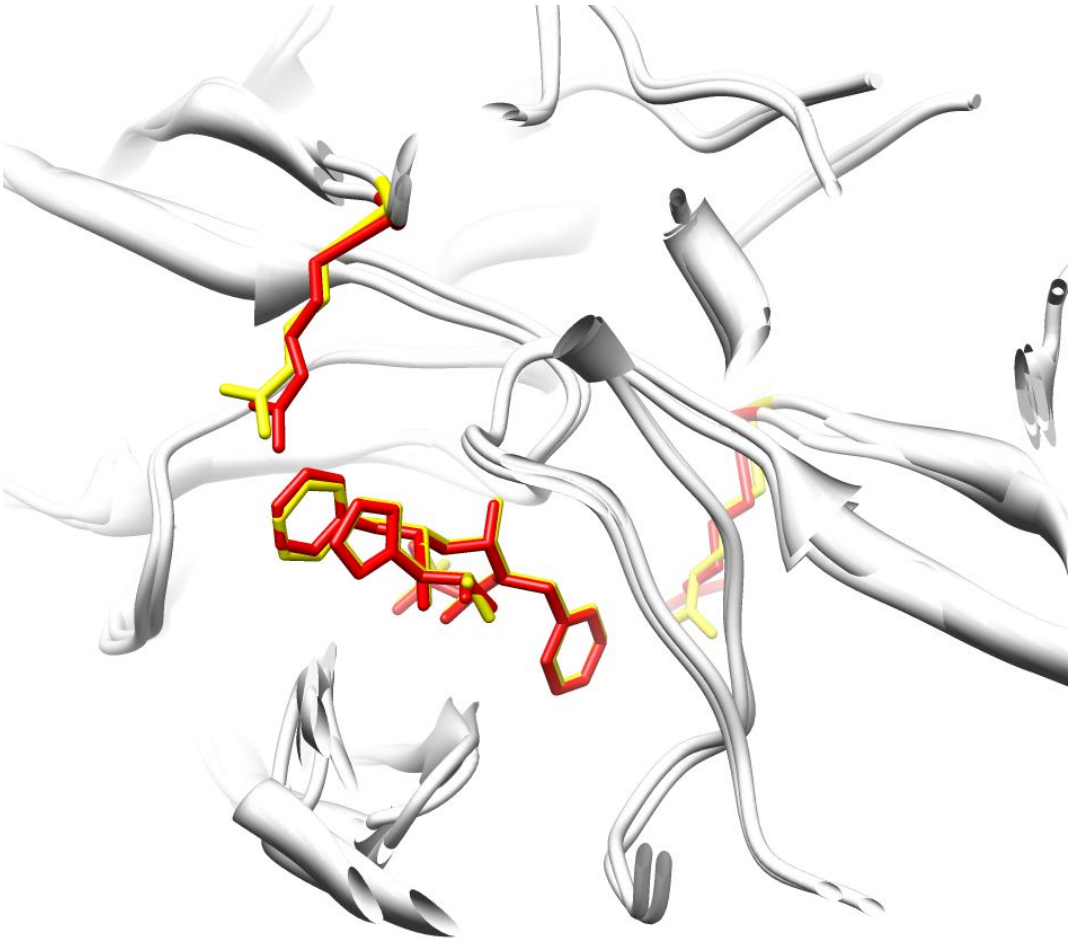
o. Predicted pocket by GalaxySite

p. Predicted pocket by ICMPocketFinder

3.3.1. Test results on the HIV protease set

The HIV protease benchmark set has 20 protein-ligand complexes, and cross-docking tests were therefore performed on 20 × 20 = 400 cases. According to Table 3.2, flexible docking with GalaxyDock improves the success rate (98.8% and 98.0% for 2 and 4 flexible residues, respectively) and average ligand RMSD (1.03 Å and 1.02 Å for both 2 and 4 flexible residues, respectively) over those obtained by rigid docking with LigDockCSA (success rate of 75.0% and average RMSD of 2.76 Å). GalaxyDock and LigDockCSA results for individual test cases are presented in Supplementary table S3. It has been previously shown that ensemble docking with AutoDock3 improved the success rate from 72.0% of rigid docking to 87.0% (Österberg *et al.,* 2002). FLIPDock showed success rates of 96.3% and 93.5% when 2 and 8 residues, respectively, were treated as flexible (Zhao and Sanner, 2007). Considering that the size of the docking box used in the FLIPDock tests was 53 times smaller than that used here, the CSA sampling of GalaxyDock combined with the modified AutoDock3 energy can be considered more effective for this test set. Figure 3.2 illustrates a successful example in which docking of the ligand of 1SBG to the protein structure of 5HVP resulted in accurate reproduction of the native bound structure with the ligand RMSD of 0.29 Å.

Figure 3.2. A successful example from the HIV proteases cross-docking experiment. The ligand of 1SBG was docked to the protein structure of 5HVP with two flexible residues (RMSD = 0.29 Å). The predicted ligand pose and receptor side-chain (red) agree very well with the crystal pose (yellow).



To evaluate whether the protein conformational changes are well reproduced, the prediction accuracy of the side-chain χ_1 angle of flexible residues were examined (shown in Supplementary table S4). The predicted χ_1 angle is considered accurate if its value is within 30° from the native χ_1 angle (Yanover *et al.*, 2008). The average χ_1 accuracy is 83.8% and 62.3% when 2 (ARG8 of the 2 chains) and 4 (2 ARG8s and 2 ILE50s) residues, respectively, of HIV protease are considered flexible. The lower accuracy in the case of 4 flexible residues is mainly due to inaccurate prediction of the ILE50 χ_1 angle. However, even when the ILE50 χ_1 angle is not predicted precisely, the overall hydrophobic interactions between ILE50 and ligand tend to be retained, resulting inaccurate binding pose prediction (See Figure 3.3 (a) and (b)).

3.3.2. Test results on the LXRβ set

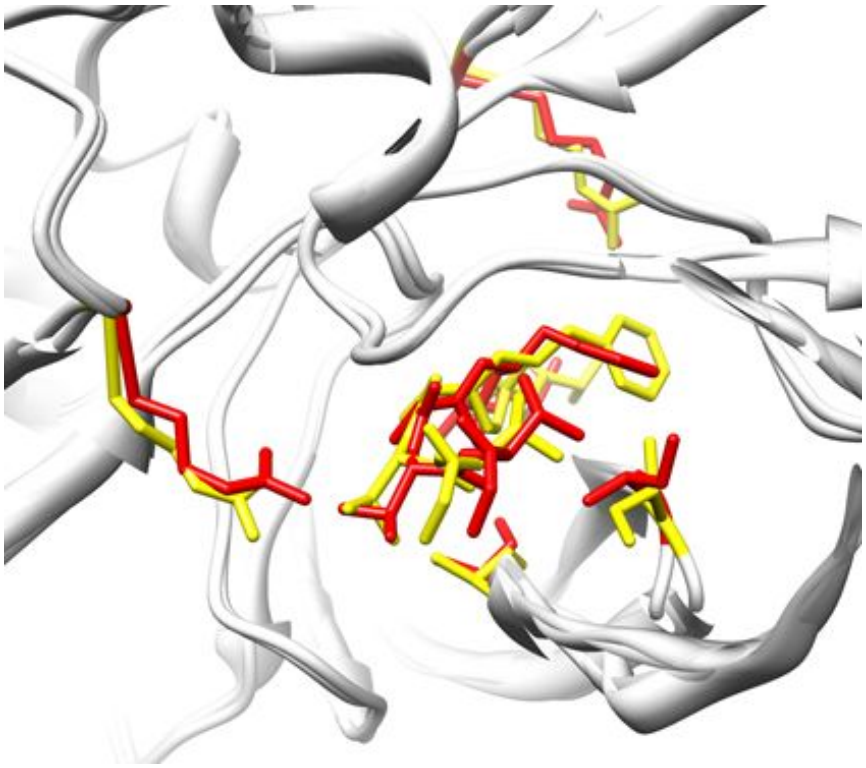
The LXRβ benchmark set consists of 3 protein-ligand complexes, and 9 cross-docking calculations were therefore performed. As shown in Table 3.2, GalaxyDock with a flexible PHE triad in the binding pocket showed a much more improved performance (success rate, 88.9%; average ligand RMSD, 1.63 Å) compared to rigid docking by LigDockCSA (success rate, 22.2%; average ligand RMSD, 3.54 Å) and by AutoDock4 (success rate, 22.2%; average ligand RMSD, 3.14 Å; Davis and Baker, 2009) (See supplementary table S5 for details). Even in the worst prediction by GalaxyDock, key interactions between flexible residues and ligand such as π-π stacking are preserved. (See

Figure 3.3 (c))

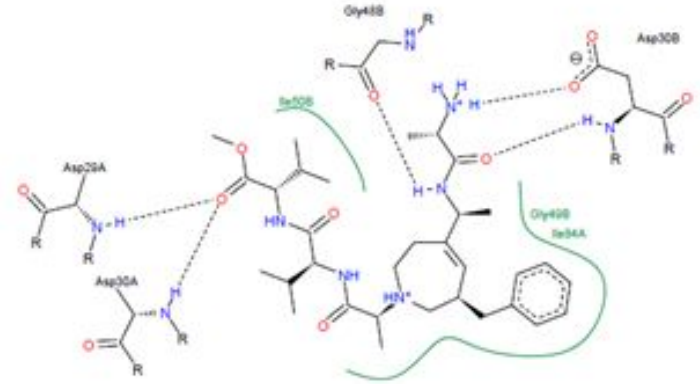
Flexible docking results by RosettaLigand are available for this set and are shown together in Table 3.2. RosettaLigand shows improved performance compared to rigid docking, but a direct comparison of its performance with that of GalaxyDock is not appropriate since the flexible degrees of freedom are set differently. Flexible residues were not pre-selected in the RosettaLigand calculations, and even backbone flexibilities were considered partially in one of the 2 tests (second row of RosettaLigand results in Table 3.2). Although the protein conformational search space of RosettaLigand is larger than that of GalaxyDock, the ligand translation and rotational space is smaller than that of GalaxyDock. The computation time of RosettaLigand is 40–80 CPU h, but that of GalaxyDock is 0.5 CPU h for this set. RosettaDock employs a Monte Carlo with minimization (MCM) method for conformational search, whereas GalaxyDock uses a CSA method, which was reported to be more efficient than MCM for diverse optimization problems (Lee *et al.*, 1997). This result implies that the CSA technique may be further exploited to incorporate backbone flexibilities in the future.

Figure 3.3. Predicted (red) and native (yellow) binding poses of HIV protease (initial protein structure from 1HPS, ligand from 1HBV, ligand RMSD = 1.32 Å, χ_1 accuracy = 75%) are shown in (a). From these poses, the same protein-ligand interaction map, shown in (b), was obtained using PoseView (Steinland *et al.*, 2007), although there are a few errors in the predicted side-chain conformations. (c) An example of LXRβ cross-docking experiments. When the ligand of 1PQC was docked to 1P8D (predicted: red and cyan, native: yellow and green, ligand RMSD: 1.97 Å, and χ_1 accuracy: 66.7%), some π-π interaction was preserved although the predicted pose has some error.

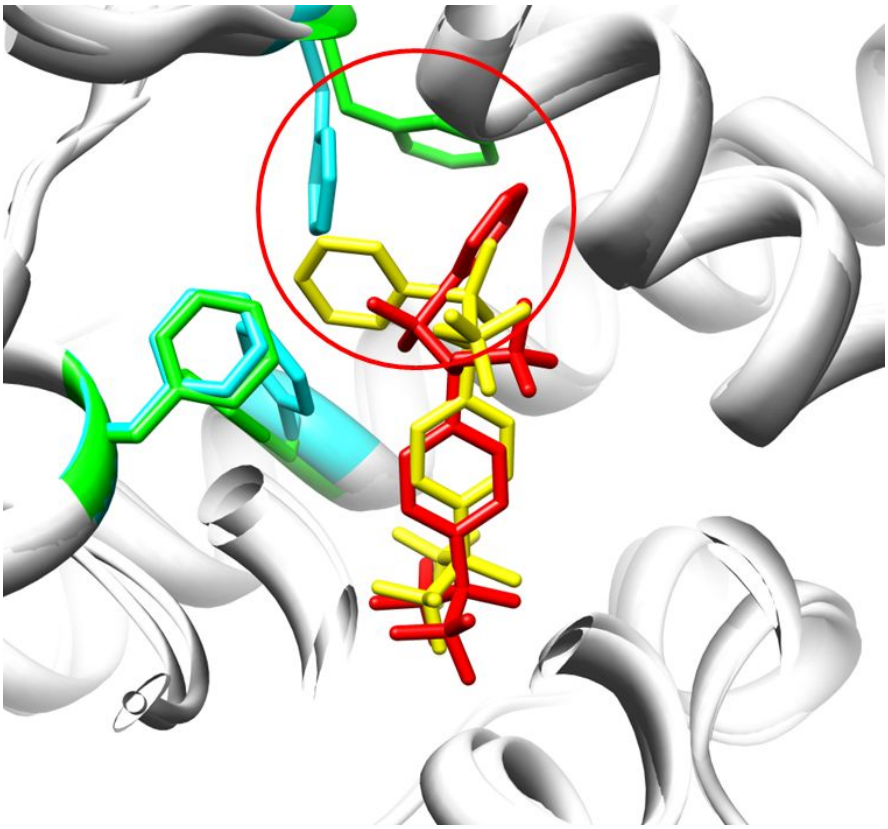
(a)



(b)



(c)



3.3.3. Test results on the cAPK set

Four crystal complex structures comprise the cAPK set, and 16 cross-docking experiments were therefore carried out. According to Table 3.2, GalaxyDock with 2 flexible residues showed a success rate of 62.5% and an average RMSD of 2.07 Å, an improvement over those of rigid docking by LigDockCSA (43.8% and 3.89 Å). Details are shown in supplementary table S6. Although cross-docking tests for the same set were performed with a modified version of UCSF DOCK which employs ITScore (Huang and Zou, 2007), direct comparison with the current results is not appropriate because the rigid docking results on the top scoring poses are not available (62.5% and 4.6 Å when the best pose is selected from the top 5 poses by energy), and the cognate protein structure is include in ensemble docking (100% and 0.35 Å). We were unable to identify the results of other flexible docking methods that sample protein with ligand conformation simultaneously, except for the FLIPDock (Zhao and Sanner, 2008) cross-docking results for 3 cases of cAPK. In 2 cases (1STC ligand docked to protein structures of 1BKX and 1YDT), FLIPDock shows better results (0.85 Å and 0.56 Å) than GalaxyDock (2.52 Å and 2.51 Å), while in docking of the 1YDT ligand to the 1BKX protein structure, GalaxyDock shows better performance (1.29 Å) than FLIPDock (6.20 Å).

The worse performance for this test set seems to be due to the more flexible backbone structure. Previous study (Wong *et al.*, 2005) have indicated that the binding pocket of cAPK involves hinge-like displacements in the GLY-rich flap, from GLY50 to VAL57, and

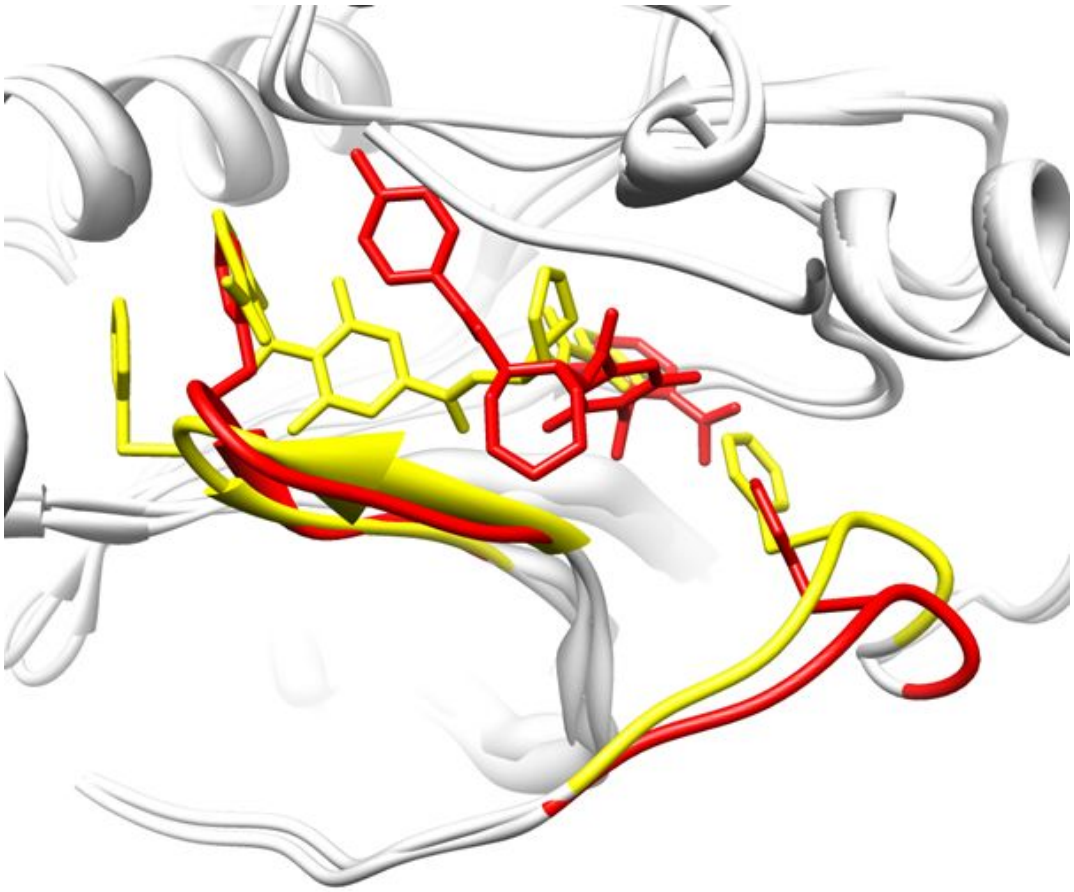
conformational changes of a flexible loop, from THR324 to GLU331. Allowing only side-chain flexibility of 2 residues, PHE54 in the GLY-rich flap, and PHE327 in the flexible loop, prohibits placement of a large ligand in the correct position when the volume of the binding site undergoes a large change, as shown in Figure 3.4. This case clearly illustrates a possible limitation of the current version of GalaxyDock, which allows only side-chain flexibility of pre-selected residues. However, such limitation could be partially overcome in the future by combining with backbone sampling by a modern loop modeling method (Park and Seok, 2012, Ko *et al.*, 2011).

3.3.4. Test results on the diverse set

The diverse set consists of 32 crystal structures involving 16 proteins. A total of 30 cross-docking experiments were performed (2 cross-docking experiments for 14 proteins and 1 docking to the apo structure for aldose reductase and anti-steroid fab). The number of flexible residues varied from one to five as shown in Table 3.1. As shown in Table 3.2, success rates of GalaxyDock (86.7% for known pocket and 80% for predicted pocket) are much higher than those of LigDockCSA rigid docking (43.3% and 46.7%). Details of docking result is shown in supplementary table S7. Higher performance of LigDockCSA by 3.4% for predicted pocket corresponds to 1 cross-docking out of 30, so it may not be statistically significant. The performance of GalaxyDock is comparable to that of SCARE (success rates of 90.0% for known binding pocket and 80.0% for predicted

pocket) although the results for predicted pocket contain the effect of the two different binding site prediction methods. SCARE has an advantage in that it does not require pre-knowledge of flexible residues. However, if flexible residues can be predicted reliably, GalaxyDock would also be a useful method because it is 60 times faster than SCARE (The average computation time is ~15 h for SCARE and ~15 min for GalaxyDock for this set.).

Figure 3.4. Conformational changes (from red, 1STC, to yellow, 1BX6) in the GLY-rich flap (on the left) and the variable loop (on the right) prohibit accurate placement of the ligand of 1BX6 (predicted: red, native: yellow) starting from the protein structure of 1STC.

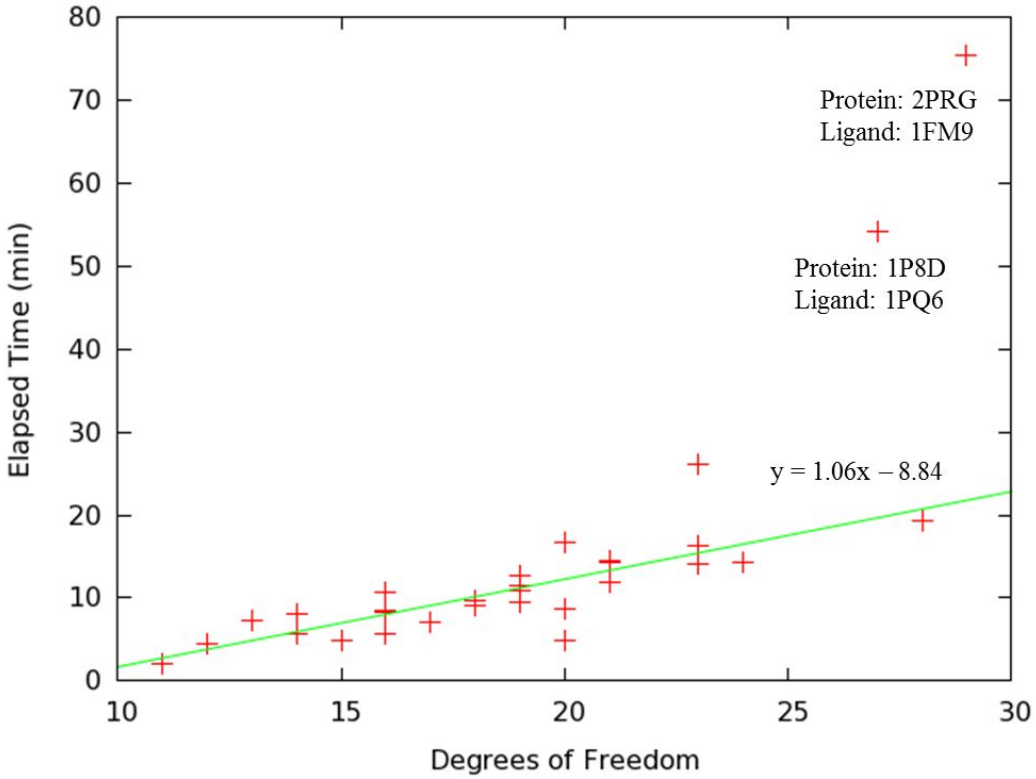


The computation time increases mildly with the number of degrees of freedom N_{dof} with two outliers at large N_{dof} of 29 and 27 (see Figure 3.5). Although the current set does not contain enough number of examples for larger degrees of freedom, it may be said that GalaxyDock works efficiently when $N_{dof} < 25$ or the maximum number of flexible residues < 5.

3.3.5. Effect of using rotamer library

GalaxyDock samples side-chain conformations from the rotamer library in the initial bank generation stage. The side-chain conformations may deviate from the rotamer conformations during subsequent energy minimizations carried out after conformation mixing and perturbation during CSA global optimization. For example, in the cross docking of Ligand 1PMN-Receptor 1PMV in the diverse set, χ_2 angle of residue M146 in the top scoring pose shows deviation of 52 degrees from the closest rotamer but is only five degrees from the experimental structure. However, when flexible residues take angles of extremely low probabilities in the rotamer library, the current strategy shows limitation. For example, when the ligand of 9HVP binds to HIV protease, the χ_1 angles of ARG8 of chain A and B in the rotamer library closest to the native angles have very low probabilities of 0.07% and 0.14%, respectively. A similar problem of non-rotamericity was also identified in previous studies (Heringa *et al.*, 1999, Zavodsky *et al.*, 2005, Jackson *et al.*, 1999). Finding an improved side-chain sampling strategy is therefore an important future goal.

Figure 3.5. Correlation plot of the diverse set using experimental pocket between degrees of freedom and time. The green line is a trend line except two outliers.



3.4. Conclusion of this section

Incorporating protein flexibility is an important issue in protein-ligand docking, but flexible docking accompanies the risk of increasing false positives due to the enlarged conformational search space. In this study, we introduce a new flexible protein-ligand docking program called GalaxyDock, as our first attempt to redress this problem by allowing side-chain flexibility of pre-selected protein residues. The relatively straightforward strategy of extending the global optimization algorithm CSA (Shin *et al.*, 2011) and the AutoDock-based energy function from our previous rigid docking program LigDockCSA appears to be successful, increasing the success rate by 10%–60% and reducing the average ligand RMSD from native by 50%. GalaxyDock also showed better or comparable results when compared with other flexible docking methods for 4 test sets. Encouraged by these results, local backbone flexibility may be further incorporated by simultaneous optimization of local protein loop structure and ligand conformation. Recent developments in protein loop modeling may be utilized for this purpose (Park and Seok, 2012, Ko *et al.*, 2011), and incorporation of larger protein motions would be the next step. Automatic detection of flexible regions is also required for more practical applications, and a model consensus method tested in homology modeling area may be employed in the future (Park *et al.*, 2011). An important extension of the current work would be to contribute to more accurate estimation of binding affinity by flexible docking, and such study is underway.

4. GalaxyDock2: improving GalaxyDock using beta-complex and binding affinity prediction

4.1. Overview of this section

Up to this point, a rigid receptor docking program, LigDockCSA, and a flexible receptor docking program, GalaxyDock, were introduced. They incorporates a powerful optimization technique called CSA with scoring functions mainly based on AutoDock3 scoring function (Morris *et al.*, 1999). Even though they have shown successful benchmark results, there remains a possibility for improvements; quality of initial bank and binding affinity prediction.

In this section, a new version of GalaxyDock called GalaxyDock2 is introduced. It shows significantly improved performance both in binding mode and binding affinity prediction compared to the previous version (now called GalaxyDock1). While maintaining the general CSA global optimization protocol, a more effective method for the generation of initial conformations is used in GalaxyDock2. More specifically, a fast geometry-based docking method that employs the beta-complex, (Kim *et al.*, 2010, 2011) a structure derived from the Voronoi diagram of receptor atoms, is used to generate the initial set of conformations called the initial “bank” in CSA. The initial bank in CSA plays an essential role because it provides a source of conformational diversity in subsequent iteration steps as well as serving

as the initial starting point for CSA iteration. This new feature could enhance both the computational speed and binding mode prediction accuracy of global optimization.

GalaxyDock2 also provides a new binding affinity function, while the same energy function employed in GalaxyDock1 is used to guide conformational sampling. The binding affinity function is composed of the protein-ligand interaction energy and the ligand torsion angle energy from the GalaxyDock energy function. The ligand torsion angle energy in the unbound state is also included. The binding affinity function was obtained by three-fold cross validation on the PDBBind set (Cheng *et al.*, 2009), which consists of 195 complexes of known binding affinities, and tested on a separate set of 80 complexes from AffinDB (Block *et al.*, 2007).

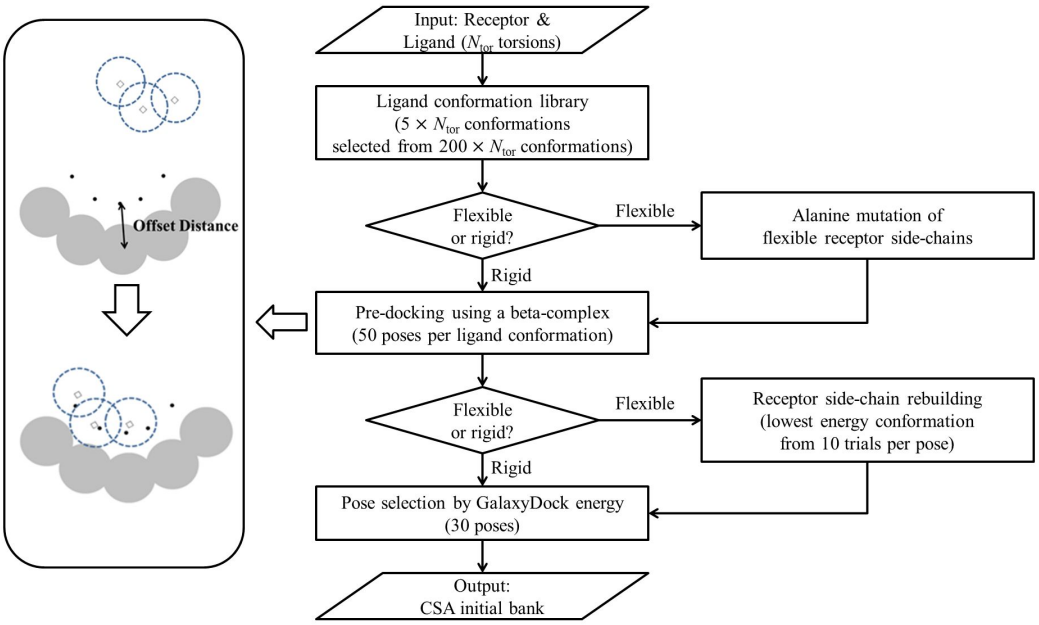
Finally, possible applications of GalaxyDock2 in virtual screening were investigated on four proteins in terms of the following three measures: enrichment factor (EF), true positive rate at a low false positive rate, and the Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC) measure. BEDROC is related to area under receiver operating characteristic curve (AU-ROC), but emphasizes early recognitions (Truchon and Bayly, 2007).

4.2. Methods

4.2.1. Initial bank generation using Voronoi diagrams

To improve the quality of the initial bank for CSA global optimization, the theory of beta-complex (Kim *et al.*, 2010, 2011) was employed in the "pre-docking" stage of initial bank generation. Beta-complexes are derived from Voronoi diagrams of protein atoms and provide geometric representations of protein surfaces. Details about the docking method using beta-complexes can be found in Kim *et al.* (2011). To summarize, pre-docking using beta-complexes is carried out by minimizing the distances between the coordinates of ligand atoms (at a fixed ligand conformation) and those of the same number of offset points generated from the protein surface, represented by a beta-complex. The offset points are generated by moving tangent points to the receptor surface atoms by an "offset distance", and a subset of consecutive offset points are selected randomly for pre-docking. A schematic illustration of this pre-docking method is given in Figure 4.1, together with a flowchart of the overall initial bank generation phase.

Figure 4.1. Flowchart of the procedure of initial bank generation. The pre-docking phase is illustrated to the left. Docking of a ligand (blue dashed circles) to the protein surface represented by a beta-complex (gray circles) was performed by minimizing the distances between centers of ligand atoms and offset points (black dots).



Since the pre-docking stage is a rigid-ligand docking process, ligand flexibility is ascertained by performing pre-docking for a pool of ligand conformations. To assure conformational diversity, the conformation pool is generated as follows: (1) random values are assigned to free ligand torsion angles to generate $200 \times N_{tor}$ conformations, where N_{tor} is the number of free ligand torsion angles; and (2) $5 \times N_{tor}$ conformations are then selected by a max-sum algorithm. In the max-sum algorithm, the lowest-energy conformation is selected first, and the structure with the largest root-mean-square-distance (RMSD) from the first one is selected next. In the following steps, the conformation that is most dissimilar from the previously selected structures is selected. The sum of the RMSDs from the previously selected conformations is used as the measure of dissimilarity (Snarey *et al.*, 1997). For step (1), in addition to uniform sampling of ligand torsion angles, random sampling was implemented. However, random sampling tended to generate more diverse ligand conformations, and the initial bank resulted in higher CSA performance.

When the receptor protein is held rigid, 50 docking poses are generated in the pre-docking stage for each of the $5 \times N_{tor}$ conformations, and 30 lowest-energy poses are selected as the initial set of conformations that constitute the CSA initial bank. When side-chain flexibility of the receptor is taken into account, pre-docking is performed on the receptor structure with flexible side chains mutated to alanine to generate $50 \times 5 \times N_{tor}$ docking poses. The CSA initial bank is formed by selecting the 30 lowest energy poses after recovery of the

mutated side-chains by selecting the lowest receptor-ligand conformation out of 10 trial conformations generated by randomly attaching side-chains from the Dunbrack backbone-dependent rotamer library (Dunbrack, 2002). We use conformations selected from a small number of randomly chosen rotamers (e.g., 10 rotamers were used in this case) because more extensive rotamer sampling on the initial ligand poses, which are not always accurate enough, could introduce wrong bias in the side-chain conformations in the initial bank.

The two parameters involved in the pre-docking stage are "β value" used to define the beta-shape of the protein and "offset distance" to locate the offset points. A β value of 1.2 Å and an offset distance of 1.8 Å were chosen by testing on the ASTEX diverse set (Hartshorn *et al.*, 2007). Details provided in Supporting Information Table S8.

4.2.2. Benchmark test sets for binding mode prediction

For the benchmark testing of rigid-receptor docking, the ASTEX diverse set consisting of 85 high-quality protein-ligand crystal structures was employed because test results for this set using other docking programs such as AutoDock3, GOLD, and Surflex-Dock are available in the literature. For the performance testing of flexible-receptor docking, the following three sets employed in previous section are used: liver X receptor beta (LXRβ), cAMP-dependent protein kinase (cAPK), and a diverse set of 16 proteins. The set for LXRβ and the diverse set were selected because test results for both

are available in the literature: RosettaLigand and AutoDock4 on LXRβ, and SCARE on the diverse set. The cAPK set was used to illustrate a case involving backbone structural change. The previously used HIV protease set was excluded because it is a relatively easy set. The binding pocket definition and selected flexible residues were kept the same as in Section 2 and 3.

4.2.3. Development of binding affinity function

The energy function of GalaxyDock is composed of the following three parts: protein-ligand interaction energy ($E_{prot\text{-}lig,AD}$), calculated based on the AutoDock3 scoring function (Morris *et al.*, 1999); ligand internal energy from AutoDock3 ($E_{lig\text{-}int,AD}$) and ligand torsion energy from Piecewise Linear Potential (PLP; $E_{lig\text{-}tor,PLP}$) (Gehlharr *et al.*, 1995); and protein internal energy estimated by the sum of the AutoDock3 van der Waals energy ($E_{prot,AD}$) and the ROTA score ($E_{prot,ROTA}$), a statistical atom-pair potential derived from the PDB (Hartmann *et al.*, 2007), when protein side-chain flexibility is taken into consideration. Mathematical expressions for each energy term are provided in Table 4.1.

Table 4.1. Scoring functions of GalaxyDock, AutoDock, and UCSF DOCK

GalaxyDock energy components

$E_{\text{prot-lig,AD}}$[a]	$w_{vdW}\sum_{i(prot)}\sum_{j(ligand)}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+w_{hbond}\sum_{i(prot)}\sum_{j(ligand)}h(t_{ij})\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right)$ $+w_{elec}\sum_{i(prot)}\sum_{j(ligand)}\frac{q_i q_j}{\epsilon(r_{ij})r_{ij}}+w_{solv}\sum_{i(prot)}\sum_{j(ligand)}(S_i V_j+S_j V_i)\exp\left(-\frac{r_{ij}}{2\sigma^2}\right)$
$E_{\text{lig-int,AD}}$[a]	$w_{vdW}\sum_{i(ligand)}\sum_{j(ligand)}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)$
$E_{\text{lig-tor,PLP}}$[b]	$w_{tor}\sum_{k=1}^{N_{rot}}F_k[1+\cos(n_k\phi_k-\phi_{ko})]$
$E_{\text{prot,ROTA}}$[c]	$w_{ROTA}\sum_{i(prot)}\sum_{j(prot)}s_{ij,ROTA}(r_{ij})$
$E_{\text{prot,AD}}$[c]	$w_{vdW,prot}\sum_{i(prot)}\sum_{j(prot)}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)$

a. The energy terms represent Lennard-Jones potential energy, orientation-dependent hydrogen bond energy, Coulomb electrostatic potential energy with distance-dependent dielectric constant, and solvation free energy, as explained in detail in Section 2. The weights (w_{vdw}, w_{hbond}, w_{elec}, w_{solv}) = (0.1485, 0.1146, 0.0656, 0.1771) are used following Morris *et al.* (1999).

b. Ligand torsion energy term of the PLP scoring function. The weight w_{tor} = 0.1 is used as in Section 2.

c. Energy terms to describe the protein internal energy. The weight parameters w_{ROTA} = 0.06 and $w_{vdW,prot}$ = 0.05 are used following Section 3. The ROTA energy is distance-dependent atom-pair statistical potential (Hartmann *et al.*, 2007)

Scoring functions used in virtual screening experiments

GalaxyDock (sampling)	$E_{GalaxyDock} = E_{prot-lig,AD} + E_{lig-int,AD} + E_{lig-tor,PLP} + E_{prot,ROTA} + E_{prot,AD}$
GalaxyDock Binding Affinity (Re-scoring of CSA final bank)	$BA = 0.65(E_{prot-lig,AD} + E_{lig-tor,PLP} - E_{lig-tor,PLP}^{(Unbound)})$
AutoDock	$E_{AutoDock} = E_{prot-lig,AD} + E_{lig-int,AD} + S_{conf,AD}$
UCSF DOCK	$E_{UCSFDOCK} = \sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + 332\sum_{i,j}\frac{q_i q_j}{\epsilon(r_{ij})r_{ij}}$

$S_{conf,AD}$: Conformational entropy from AutoDock3 which is proportional to the number of active ligand torsion angles (Morris *et al.*, 1999)

To achieve better binding affinity prediction than that obtained using the above energy function, the correlation between the energy components and the experimentally determined binding affinities for the PDBBind set (Cheng *et al.*, 2009) was first examined. The set consists of 65 proteins, for each of which crystal structures bound to three different ligands and experimental binding affinities are available. Among the different combinations of energy components, $E_{prot\text{-}lig,AD}$ and ($E_{prot\text{-}lig,AD}$ + $E_{lig\text{-}tor,PLP}$) gave the best correlations with the experimental binding affinity. Accounting for ligand conformational entropy, estimated by a constant times the number of active ligand torsion angles (as in AutoDock), did not improve the correlation. (Details are provided in Table 4.2.)

To further improve the accuracy of binding affinity prediction, the effect of subtracting the energy of a free ligand was also investigated. The energy of unbound ligand was estimated by optimizing the ligand in empty affinity grids using a short CSA run. Subtracting the free ligand energy from the interaction energy improved the correlation with the binding affinity, and $E_{prot\text{-}lig,AD}$ and ($E_{prot\text{-}lig,AD}$ + $E_{lig\text{-}tor,PLP}$) again showed the best correlation with this correction. (Details are shown in Table 4.2) This result agrees with the previous report of Huey *et al.* that including the unbound ligand energy improved the performance of AutoDock.

The final form of the binding affinity function is as follows:

$$BA = c(E_{prot-lig,AD} + E_{lig-tor,PLP} - E_{lig-tor,PLP}^{(Unbound)}) + d \qquad (4.1)$$

where $E_{lig-tor,PLP}^{(Unbound)}$ represents the ligand torsion energy in the unbound state. The empirical parameters *c* and *d* were determined to be 0.65 and 0 by a three-fold cross-validation on the PDBBind set. Details on parameter determination are discussed in the Results and Discussion section.

The binding affinity function of Eq. (4.1) was tested on a separate test set of 80 protein–ligand complexes. The set was extracted from AffinDB (Block *et al.,* 2006) using the two criteria (Lee *et al.*, 2009) of molecular weight < 500 Da and no hetero atoms in the binding pocket. We have excluded proteins with hetero atoms such as ions because we have not thoroughly tested energy parameters for such cases yet. However, we believe that such work should be performed for GalaxyDock2 in the near future. The complexes that belonged to the PDBBind set were excluded. When more than one crystal structures were available, the structure with the best resolution was selected.

Table 4.2. Accuracy of binding affinity prediction measured by the Pearson's correlation coefficient with the experimental binding affinity and root-mean-square error (RMSE) of the predicted binding affinity for different combinations of energy components. The PDBBind set (Cheng *et al.*, 2009) was used for this test.

Energy Components	Without Free Ligand Correction		With Free Ligand Correction	
	Correlation Coefficient	RMSE (kcal/mol)	Correlation Coefficient	RMSE (kcal/mol)
$E_{prot\text{-}lig,AD}$	0.444	4.375	0.469	7.012
$E_{prot\text{-}lig,AD}+E_{lig\text{-}int,AD}$	0.282	4.728	0.431	5.540
$E_{prot\text{-}lig,AD}+E_{lig\text{-}tor,PLP}$	0.398	4.066	0.475	5.911
$E_{prot\text{-}lig,AD}+S_{conf,AD}$	0.336	4.344	0.451	4.583
$E_{prot\text{-}lig,AD}+E_{lig\text{-}int,AD}+E_{lig\text{-}tor,PLP}$	0.232	4.793	0.427	4.640
$E_{prot\text{-}lig,AD}+E_{lig\text{-}int,AD}+S_{conf,AD}$	0.115	5.622	0.358	4.091
$E_{prot\text{-}ligAD}+E_{lig\text{-}tor,PLP}+S_{conf,AD}$	0.238	5.387	0.416	4.162
$E_{prot\text{-}lig,AD}+E_{lig\text{-}int,AD}+E_{lig\text{-}tor,PLP}+S_{conf,AD}$	0.030	6.706	0.301	4.255

$E_{prot\text{-}lig}$: Protein-ligand interaction based on AutoDock3 scoring function

$E_{lig\text{-}int}$: Ligand internal energy from AutoDock3

$E_{lig\text{-}tor}$: Torsion angle strain energy calculated by the torsion part of the PLP scoring function (Gehlharr *et al.*, 1995)

S_{conf}: Conformational entropy from AutoDock3 which is proportional to the number of active ligand torsion angles (Morris *et al.*, 1999)

4.2.4. Virtual screening benchmark set

Using the binding affinity function described above, GalaxyDock2 was applied to the virtual screening of compound libraries and compared with AutoDock4 (Morris *et al.*, 2009) and UCSF DOCK6 (Mukherjee *et al.*, 2010). The two docking programs were chosen for comparison because they use a force field-based scoring function similar to GalaxyDock.

For a benchmark set, Gilson's set (Jorissen and Gilson, 2005) was used because well-prepared compound libraries of medium size are provided. Jorissen and Gilson compiled 50 inhibitors for each of the following five proteins to test their support vector machine method: cyclin-dependent kinase 2 (CDK2), cyclooxygenase-2 (COX-2), factor Xa (FXa), phosphodiesterase-5 (PDE5), and α_{1A} adrenoceptor. They also gathered 1,892 background compounds from the National Cancer Institute diversity set. In this study, 950 compounds were selected randomly from the background set and added to the compound database for each protein. Therefore, each database contains 50 active and 950 decoy compounds.

Since GalaxyDock2 requires a receptor structure (unlike in Gilson's work), α_{1A} adrenoceptor, whose structure is not available in RCSB PDB, was removed from our benchmark set. The PDB IDs of the selected receptor structures for the four remaining proteins are 4GCJ (CDK2, resolution = 1.42 Å), 3NT1 (COX-2, resolution = 1.73 Å), 2PR3 (FXa, resolution = 1.50 Å), and 3TGG (PDE5, resolution = 1.91 Å).

4.2.4.1. Virtual screening using GalaxyDock2

GalaxyDock2 was run both in the rigid-receptor and the flexible-receptor mode. A cubic docking box of side length 22.5 Å was positioned at the geometrical center of the ligand in the crystal structure. All dimensions of the docking box were divided into a grid of 61 × 61 × 61 points with a grid spacing of 0.375 Å. Further details of CSA global optimization are described in Section 2 for rigid-receptor docking and in Section 3 for flexible-receptor docking. The bank size was set to 30, although a larger bank size was used in Section 2. The binding affinity was estimated by applying Eq. (1) to the top scoring binding pose obtained by CSA.

The flexible residues selected for the flexible-receptor docking are L10 for CDK2, F85 for COX-2, Y99 and Q192 for FXa, and Q817 for PDE5. These residues were selected because they show steric contacts upon superposition of available crystal structures bound to different ligands. Considering a small number of flexible residues may be justified by a previous report in which 85% of the complexes changed side-chain rotamer states in less than three residues (Gaudreault *et al.*, 2012). In addition, in GalaxyDock, considering more than four flexible residues will require significantly more conformational sampling according to previous work. (See Section 3)

4.2.4.2. Virtual screening using AutoDock4

The docking box and grid points were kept the same as in the GalaxyDock2 runs. Ten docking runs using the Lamarckian genetic

algorithm were performed for each compound. The population size was set to 150. The maximum number of generations = 27,000 and the maximum number of energy evaluations = 2.5×10^6.

4.2.4.3. Virtual screening using UCSF DOCK6

In the sphere-generation process of UCSF DOCK6, the Connolly surface (Connolly, 1993) of the receptor was generated using a 1.4 Å-radius probe, and spheres were then created on the molecular surface producing approximately one sphere per surface point. To define the docking box, spheres within 10 Å from the crystal ligand position weres elected, and extra margins of 8 Å from the selected spheres. An energy grid with a spacing of 0.3 Å was generated using the all-atom model, with an energy cutoff distance of 10 Å and distance-dependent dielectric constant ε = 4r. "Maximum orientations per ligand" and "configurations per cycle" in the ligand growth process were set to 500 and 10, respectively.

4.2.4.4. Measures for assessing virtual screening results

To assess virtual screening (VS) performance enrichment factor (EF) (Kokh *et al.*, 2008), true positive rate (TPR) at 1% false positive rate (FPR), and Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC) (Truchon and Bayly, 2007) were used.

EF of the top τ% subset of the database is calculated as

$$EF = \frac{HITS_{\tau} / N_{\tau}}{HITS_{Database} / N_{Database}} \quad (4.2)$$

where "HITS" represents the number of known active compounds in the top τ% subset or in the database, and *N* represents the total number of molecules in the specified set. A random selection of a subset from the database corresponds to EF = 1. In this work, EF was examined at 1%, 2%, 5%, and 10% subsets of the database.

In VS studies, it is important to rank hits as high as possible. However, a widely used measure, area under receiver operating characteristic curve (AU-ROC) (Klon *et al.*, 2004, Triballesau *et al.*, 2005), depends more on average ranking; so, the "early recognition problem" exists (Truchon and Bayly, 2007, Kokh and Wenzel, 2008). Two measures, TPR at 1% FPR and BEDROC are introduced to resolve this problem. The measure BEDROC adds exponential weights to early recognition as follows (Truchon and Bayly, 2007, Kokh and Wenzel, 2008):

$$BEDROC = \frac{\int_0^N F_a(x)w(x)}{\int_0^N w(x)} \tag{4.3}$$

where N is the number of compounds in database; $F_a(x)$ is the number of active compounds in the subset of high-ranked x compounds; and $\omega(x) = \exp(-\alpha x)$ is a weight factor that emphasizes early recognition, where $\alpha = 20$ as suggested in Truchon and Bayly. In actual calculations of BEDROC, the following approximate equation applicable to a discrete set of compounds was used:

$$BEDROC = \frac{\sum_{i=1}^{n} e^{-\alpha r_i/N}}{R_a\left(\frac{1-e^{-\alpha}}{e^{\alpha/N}-1}\right)} \times \frac{R_a \sinh\left(\frac{\alpha}{2}\right)}{\cosh\left(\frac{\alpha}{2}\right) - \cosh\left(\frac{\alpha}{2} - aR_a\right)} + \frac{1}{1-e^{\alpha(1-R_a)}} \quad (4.4)$$

where n is the number of hits, R_a is the number of hits in the database divided by N, r_i is the ranking of the ith hit.

4.2.5. Protein and ligand preparation

All receptor and ligand structures used in this work were prepared using Sybyl8.1. The atomic coordinates for receptors were extracted from RCSB PDB. Gasteiger–Hückel charges were assigned after attaching hydrogen atoms. For native pose reconstruction and binding affinity prediction tests, ligands separated from the PDB structures were used. Gasteiger–Hückel charges were assigned to those ligands after attaching hydrogen atoms. For VS tests, hydrogen atoms and partial charges of ligands were kept the same as in the Gilson's database (Jorissen and Gilson, 2005).

4.3. Results and discussion

4.3.1. Binding mode prediction

GalaxyDock performs searches of protein and ligand conformational degrees of freedom using conformational space annealing (CSA) (Lee *et al.*, 1997)—a population-based global optimization technique. Performances of population-based methods tend to depend on

the quality of the initial population (Reeves, 1995, Park *et al.*, 2003). The previous version, GalaxyDock1, also depended on the initial pool of conformations (called "initial bank" in CSA; data not shown). GalaxyDock1 generates the initial bank by energy minimization, after assigning random values to the ligand degrees of freedom (translation, rotation, and torsions) and randomly chosen rotamers (Dunbrack, 2002) to flexible protein side-chains. Since the initial bank in CSA is a major source of conformational diversity as well as a starting point of conformational evolution, it is highly expected that the performance of GalaxyDock can be improved by improving the quality of the initial bank.

GalaxyDock2 adopts a fast geometry-based docking method that employs a beta-complex (Kim *et al.*, 2010, 2011), derived from the Voronoi diagram to describe the shape of the protein surface, to produce a higher-quality initial bank more efficiently. A high-quality CSA initial bank is required to have enough conformational diversity because subsequent conformational searches are dependent on conformational mixing with the initial bank. It is also highly promising if the initial bank contains conformations with native-like elements in individual degrees of freedom. Random generation of the initial bank—as was done for GalaxyDock1—results in a relatively low probability of successful positioning of ligands in the binding pocket, requiring a large number of trials, especially for tight pockets. Pre-docking using beta-complex shows a much higher probability of proper ligand positioning, with no steric clashes. Because this method is based on

purely geometric considerations, not requiring energy-based optimization, fast computation is also possible.

In the case of rigid docking, GalaxyDock2 with a bank size N_{bank} = 30 produces slightly better results than GalaxyDock1 with N_{bank} = 100, when tested on the 85 protein–ligand complexes of the ASTEX diverse set (See Table 4.3). According to our previous work, GalaxyDock1 (with N_{bank} = 100) performed better than AutoDock3, GOLD (Hartshorn *et al.*, 2007), and Surflex-Dock (Spitzer and Jain, 2012) on the same set, as shown in Table 4.3. A successful example is illustrated in Figure 4.2 (a). When compared to GalaxyDock1 run with a smaller bank size of N_{bank} = 30, GalaxyDock2 (with N_{bank} = 30) shows improvement in the success rate by 5.9% and in average RMSD by 0.24 Å. The success rate is defined as the percentage of the cases in which RMSD of the top ranking pose from the native binding mode is less than 2 Å, and the average RMSD refers to the average of the RMSDs of the top ranking poses over the test set complexes.

This result shows that the new initial bank generation method is quite successful, making it possible to use a much smaller bank size without loss of accuracy. An example is illustrated in Figure 4.3. The minimum RMSD in the initial bank for 1R1H generated using the beta complex is smaller, by 5 Å, than that generated by the random sampling employed in GalaxyDock1, as shown in Figure 3 (a). This enhancement in the initial bank is transferred to the final bank obtained by CSA, leading to improvement in the predicted RMSD (RMSD of the lowest energy pose) by 2 Å, as can be seen in Figure 4.3 (b).

Table 4.3. Performance of GalaxyDock in the rigid-receptor docking mode, tested on the ASTEX diverse set (Hartshorn *et al,* 2007).

Docking Program	Success Rate[a] (%)	Average RMSD[b] (Å)
GalaxyDock2 (N_{bank} = 30)	85.9	1.24
GalaxyDock1 (N_{bank} = 30)	80.0	1.48
GalaxyDock1 (N_{bank} = 100)[a]	84.7	1.34
AutoDock3[a]	81.7	1.60
GOLD[b]	80.5	-
Surflex-Dock[c]	80.0	1.66

a. Percentage of the cases in which RMSD of the top scoring pose from that of the crystal structure is less than 2 Å

b. RMSD of the top scoring pose from that of the crystal structure averaged over the targets in the set

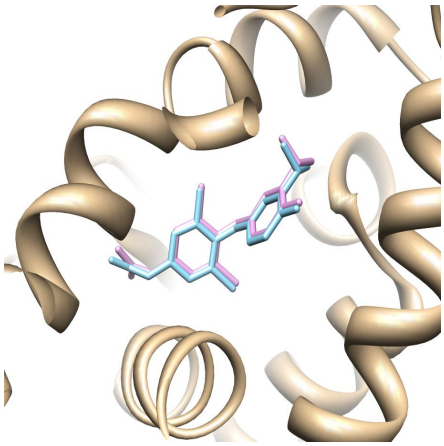
c. Taken from Section 2

d. Taken from Hartshorn *et al.*. Average RMSD is not reported in the reference.

e. Taken from Spitzer and Jain

Figure 4.2. Successful docking examples of GalaxyDock2. (a) Rigid-receptor docking result of the cognate ligand for the complex 1NAV from the ASTEX set. Root-mean-square-distance (RMSD) of the top scoring predicted pose (pink) from the crystal structure binding mode (sky blue) is 0.39 Å. (b) Cross-docking result of the ligand of 1KI4 to the protein structure of 1KIM with two flexible residues taken from the diverse set. RMSD of the predicted pose (pink) from the crystal pose (sky blue) is 0.92 Å. Flexible residues are in the same color as the ligand.

(a)



(b)

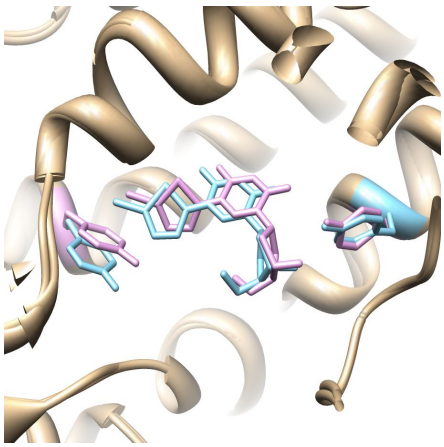
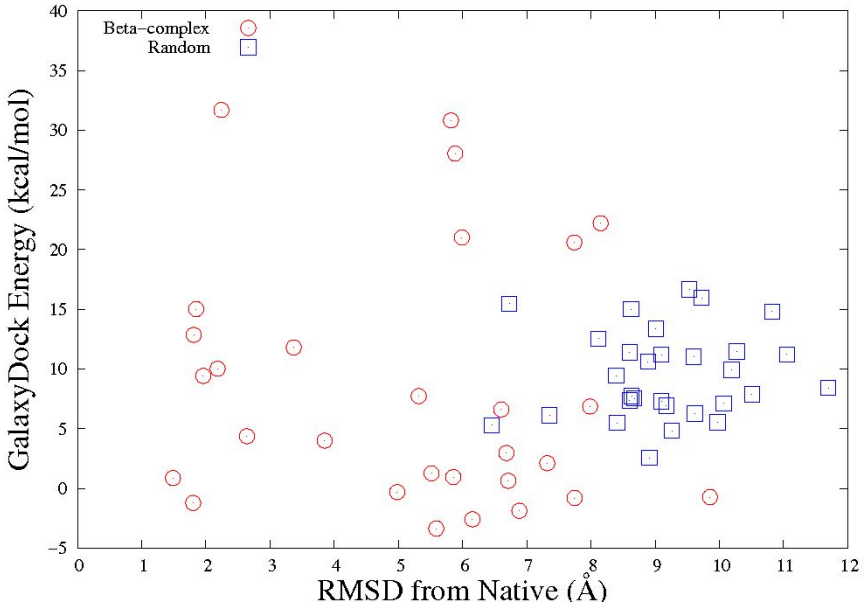
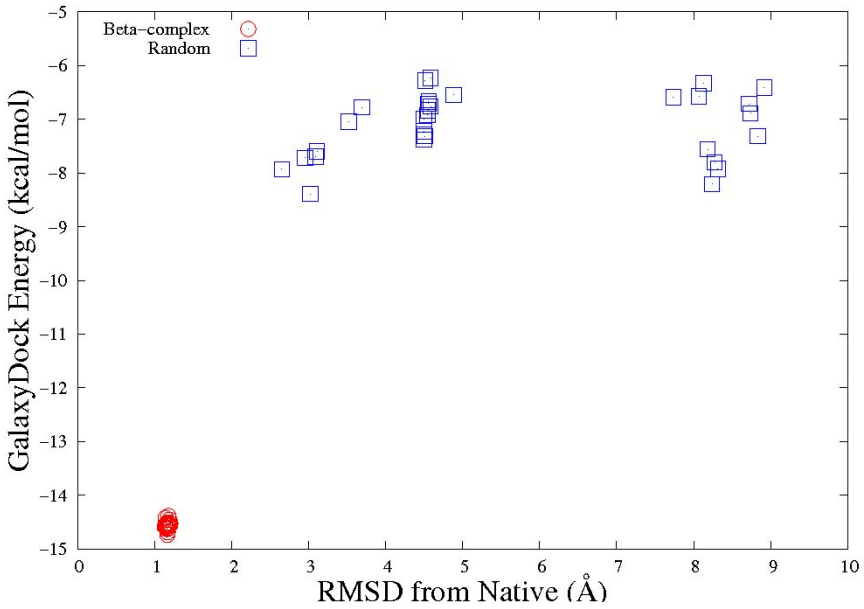


Figure 4.3. Plot of energy versus RMSD (ligand RMSD from the crystal pose) for (a) the CSA initial bank, and (b) the CSA final bank of 1R1H generated using a beta-complex (red circles) and a random sampling of GalaxyDock1 (blue squares).

(a)



(b)



Use of smaller bank size due to the improved performance reduces the global optimization effort and, thus, the computation time considerably. GalaxyDock1 with N_{bank} = 100 takes ~30 minutes, while GalaxyDock2 with N_{bank} = 30 takes ~7 minutes on a single CPU.

In the case of flexible side-chain docking, a smaller bank size of N_{bank} = 30 has already been attempted (even with GalaxyDock1) to reduce computational time, because computational complexity increases considerably in flexible-receptor docking. GalaxyDock2 with N_{bank} = 30 was tested on the same benchmark test sets: LXRβ (3 crystal structures, 9 cross-docking experiments), cAPK (4 crystal structures, 16 cross-docking experiments), and a diverse set (16 proteins and 30 cross-docking experiments). The results are summarized in Table 4.4, and a successful case is illustrated in Figure 4.2 (b)**.** The improved initial bank generation method in flexible side-chain docking also improves the docking performance of GalaxyDock2. However, the effect is relatively mild, with increased success rate (6%) for the cAPK set and the same success rates for the LXRβ and diverse sets compared to GalaxyDock1. The smaller success rate (< 70%) for the cAPK set is because of the change in the backbone structure induced by ligand docking, as described previously. Both GalaxyDock2 and GalaxyDock1 show comparable performances to other state-of-the-art flexible docking programs such as RosettaLigand (Davis and Baker, 2009) and SCARE (Bottegoni *et al.*, 2008). When predicted pockets were used for the cross-docking tests on the diverse set, GalaxyDock2 showed an improved successrate of 3% more than that of GalaxyDock1 or

SCARE, as shown in parenthesis in Table 4.4. The improvement of GalaxyDock2 over GalaxyDock1 is statistically significant with a P-value = 0.02 determined by a Student's *t*-test.

The low improvement on using the beta-complex-based initial bank generation method in flexible side-chain docking compared to rigid docking may be understood from the fact that protein flexibility is considered only indirectly in the pre-docking stage. Ligands are positioned in the alanine-mutated protein structure first, and side-chain conformations are recovered afterwards. The quality of the initial bank measured by the distribution of the GalaxyDock score and RMSD from the native mode is indeed not enhanced greatly compared to the previous initial bank generation method in flexible docking (data not shown).

Rigid docking experiments were also carried out on these flexible docking test sets, as summarized in Table 4.4. The success rates obtained by rigid docking methods are all lower than those of flexible docking methods for the current test sets, in which flexible residues are rather well defined. Among the different rigid docking methods, GalaxyDock2 shows the best performance for these sets. It remains to be investigated how the relative performance of rigid versus flexible docking methods is affected when the accurate prediction of flexible residues is unattainable.

Table 4.4. Flexible docking benchmark results: LXRβ, cAPK, and the diverse set.

Docking Program	Success Rate (%)	Average RMSD (Å)
LXRβ		
GalaxyDock2	88.9	1.54
GalaxyDock2 (rigid)	55.6	1.97
GalaxyDock1[a]	88.9	1.63
GalaxyDock1 (rigid)[a]	22.2	3.54
RosettaLigand[b]	55.6	1.91
AutoDock4 (rigid)[b]	22.2	3.14
cAPK		
GalaxyDock2	68.8	2.11
GalaxyDock2 (rigid)	43.8	2.80
GalaxyDock1[a]	62.5	2.07
GalaxyDock1 (rigid)[a]	43.8	3.89
Diverse set[d]		
GalaxyDock2	86.7 (83.3)	1.57 (1.68)
GalaxyDock2 (rigid)	53.3 (53.3)	2.19 (2.40)
GalaxyDock1[a]	86.7 (80.0)	1.70 (2.05)
GalaxyDock1 (rigid)[a]	46.7 (46.7)	3.01 (2.76)
SCARE[c]	90.0 (80.0)	N/A
SCARE (rigid) (VDWMAX = 4.0[e])[c]	40.0 (43.3)	N/A
SCARE (rigid) (VDWMAX = 1.0[f])[c]	50.0 (46.7)	N/A

a. Taken from Section 3

b. Taken from Davis and Baker

c. Taken from Bottegoni *et al.* (2008) RMSD values are not available.

d. Results for predicted pocket in parenthesis

e. The value of van der Waals potential is truncated at 4.0 kcal/mol.

f. The value of van der Waals potential is truncated at 1.0 kcal/mol.

4.3.2. Binding affinity prediction

Another key addition to GalaxyDock2 is binding affinity prediction. The energy function of GalaxyDock1 could be used to estimate the binding affinity, but more accurate prediction is made possible by taking an optimal combination of energy components and by subtracting the energy of the unbound ligand state, as described in Section 4.2.3. The best correlation with experimental binding affinity was obtained when protein-ligand interaction energy and ligand torsion energy were combined with the unbound ligand state correction, according to Table 4.2, Although the correlation coefficient improves with the free-ligand correction, the root-mean-square-error (RMSE) of predicted binding affinity increases. This is because the weight parameters of the GalaxyDock energy function were taken from the AutoDock energy function, and they were fitted to minimize prediction errors without free-ligand correction. With the current correction of the unbound ligand state, the energy function has to be re-scaled to minimize RMSE. The two empirical parameters c and d were therefore introduced in Eq. (4.1) and determined by a three-fold cross-validation on the PDBBind set (Cheng *et al.*, 2009), as summarized in Table 4.5. The parameters c and d were varied from 0.0 to 1.0 in 0.01 and 0.05 increments, respectively, for each training set; c = 0.65 and d = 0 were determined to be optimal.

Table 4.5. Three-fold cross-validation on the PDBBind set (Cheng *et al.*, 2009) to determine the parameters *c* and *d* of the binding affinity function

Set No.			Training Set		Test Set	
	c	*d*	Correlation Coefficient	RMSE (kcal/mol)	Correlation Coefficient	RMSE (kcal/mol)
1	0.65	0.0	0.462	3.390	0.511	3.198
2	0.63	0.0	0.473	3.210	0.471	3.540
3	0.63	0.0	0.489	3.358	0.442	3.251

An additional test was carried out on a separate set called the AffinDB set (Block *et al.*, 2006). The results for the PDBBind training set and AffinDB test set in terms of correlation coefficient and RMSE are summarized in Table 4.6. With the new binding affinity function, GalaxyDock2 predicts binding affinity more accurately than AutoDock3, on which the energy function is based, with improvement in the correlation coefficient by ~0.1 and RMSE by ~1 kcal/mol. The correlation coefficient increased from 0.336 (AutoDock3) to 0.475 (GalaxyDock2) for the training set and from 0.393 (AutoDock3) to 0.484 (GalaxyDock2) for the test set. RMSE also improved from 4.34 kcal/mol (AutoDock3) to 3.32 kcal/mol (GalaxyDock2) for the training set and 3.95 kcal/mol (AutoDock3) to 3.12 kcal/mol (GalaxyDock2) for the test set. This consistent improvement in binding affinity prediction implies that the current modification of the AutoDock3 energy function as implemented in GalaxyDock2 may be transferable to more general binding affinity prediction problems.

Table 4.6. Accuracy of the new binding affinity prediction implemented in GalaxyDock2 on the training set and a separate test set compared with the results of AutoDock3 scoring function (Morris *et al.*, 1999) on the same sets.

Binding Affinity Function	Correlation Coefficient	RMSE (kcal/mol)
Training set: PDBBind set (Cheng *et al.*, $N_{complex}$ = 195)		
GalaxyDock2	0.475	3.32
AutoDock3	0.336	4.34
Test set: AffinDB set (Block *et al.*, $N_{complex}$ = 80)		
GalaxyDock2	0.484	3.12
AutoDock3	0.393	3.95

When the correlation coefficient of the GalaxyDock2 binding affinity was compared to those of 17 other scoring functions on the same PDBBind set, it ranked eighth (Cheng *et al.*, 2009, Table 4.7). The correlation coefficients reported in Table 4.7 were obtained by applying the scoring functions to the crystal structures of protein-ligand complexes. However, in real prediction problems, the scoring functions have to be applied to predicted binding poses. In fact, several scoring functions such as X-score and different Sybyl scores that show higher correlation than GalaxyDock2 are used mainly for re-scoring binding poses generated by a separate docking engine. As described previously, the GalaxyDock2 binding affinity function is closely related with the energy function that guides sampling during GalaxyDock2 docking. A natural and useful follow-up of this study would be to adapt the merits of other high-performance scoring functions that were previously used only in re-scoring, to the GalaxyDock docking procedure to guide conformational search.

Table 4.7. Pearson's correlation coefficients of 17 scoring functions for the PDBBind set (Cheng *et al.*, 2009). The correlation coefficients of 16 scoring functions except GalaxyDock were taken from Cheng *et al.*.

Scoring Function	Correlation Coefficient
X-Score::HMScore	0.644
DrugScoreCSD	0.659
SYBYL::ChemScore	0.555
DS::PLP1	0.545
GOLD::ASP	0.534
SYBYL::G-Score	0.492
DS::LUDI3	0.487
GalaxyDock2	0.475
DS::LigScore2	0.464
GlideScore-XP	0.457
DS::PMF	0.445
GOLD::ChemScore	0.441
SYBYL::D-Score	0.392
DS::Jain	0.316
GOLD::GoldScore	0.295
SYBYL::PMF-Score	0.286
SYBYL::F-Score	0.216

4.3.3. Virtual screening

Since GalaxyDock2 shows improved binding mode prediction both in terms of accuracy and speed, it may be applied to the virtual screening (VS) of compound libraries, when combined with the new binding affinity function described above. The usefulness of GalaxyDock2 for VS has been tested on a set of four proteins—CDK2, COX-2, FXa, and PDE5. A compound library of 1,000 compounds containing 50 known inhibitors was used for screening.

The VS results of GalaxyDock2 are compared with those of AutoDock4 and UCSF DOCK6 in Table 4.8. Among the three rigid-receptor docking programs, GalaxyDock2 shows the highest enrichment factor (EF) values at all the top τ% (τ = 1, 2, 5, and 10) for CDK2 and COX-2. For FXa, UCSF DOCK6 shows the highest EF at top 1%, but GalaxyDock2 shows the highest EF at 2%, and AutoDock4, at 5% and 10%. For PDE5, AutoDock4 shows the highest EF at 1% and 2%, and GalaxyDock2, at 5% and 10%. The activities of top-ranked compounds in VS may be tested by in vitro experiments. It is therefore important to find hits with higher ranks in VS. When EF values at 1% and 2% are compared, those of GalaxyDock2 in the rigid mode are much higher than those of the other programs for CDK2 and COX-2. For the other two proteins, EF obtained by GalaxyDock2 is similar to those by other programs. This tendency can also be confirmed by Figure 4.4, which shows fractions of known inhibitors found in the top τ% subset for each protein. The curve for GalaxyDock2 run in the rigid mode (green dashed line) is almost

always above those of the other two rigid docking programs (blue and purple lines) at smaller τ. The BEDROC measure assigns more weights to higher-ranked active compounds, and GalaxyDock2 showed the best performance, in terms of BEDROC, except for FXa (Table 4.8). The true positive rate (TPR) at 1% false positive rate (FPR) also agrees with the above tendency: GalaxyDock2 shows the best values except for PDE5.

Although we made efforts to ensure that the head-to-head comparisons across different docking programs were fair, we acknowledge that the performance of the programs depends on the parameters that are used. Therefore, researchers who use different input parameters or protocols would obtain different performances.

GalaxyDock2 in the flexible side-chain mode was also tested with one or two flexible side chains. When compared to the results of GalaxyDock2 in the rigid mode, flexible docking shows higher EF and BEDROC values, except for CDK2 (See Table 4.8 and Figure 4.4), which is known to show loop flexibility upon ligand binding. The above results imply that the explicit treatment of binding site flexibility by GalaxyDock2 can enhance the performance of VS, when information on flexible residues is available. Kokh and Wenzel also reported that flexible docking resulted in higher EF than that seen in rigid docking.

In CDK2, it is known that a 23-residue T-loop changes secondary structure (Jeffrey *et al.*, 1995), and a Gly-rich loop interacting with the ligand also shows conformational changes during ligand binding. Obviously, such large backbone movements cannot be

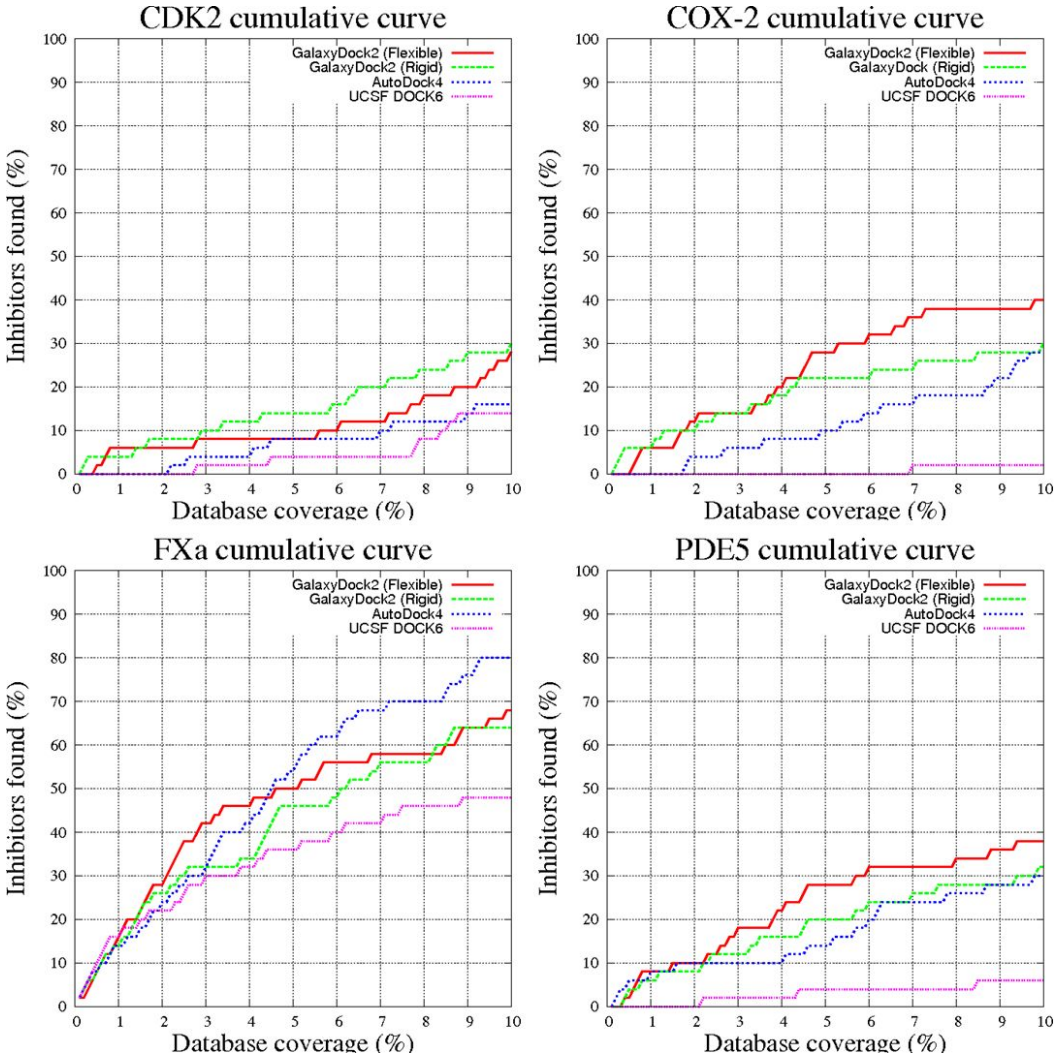
accounted for by flexible side-chain docking. GalaxyDock2, which employs a powerful global optimization technique, maybe further extended to account for flexible loops by simultaneous optimization of ligand and receptor degrees of freedom in the future. State-of-the-art loop modeling techniques (Park and Seok, 2012, Ko *et al.*, 2011) maybe useful for such development.

The average computation times per compound are: 15 seconds for UCSF DOCK6, 7 minutes for GalaxyDock2 rigid docking, 10 minutes for AutoDock4, and 13 minutes for GalaxyDock2 flexible docking. GalaxyDock2 may be unsuitable for virtual screening of compound libraries of considerably larger sizes than those considered here. However, such a problem could be solved by parallel implementations or by using hardware accelerators such as GPUs.

Table 4.8. Comparison of the results of virtual screening performed on the Gilson set (Jorissen and Gilson, 2005) for GalaxyDock2, AutoDock4, and UCSF DOCK6.

Docking Program	Enrichment Factor				TPR at 1% FPR	BEDROC
	1%	2%	5%	10%		
CDK2						
GalaxyDock2 (flex.)	6.0	3.0	1.6	2.8	6%	0.189
GalaxyDock2 (rigid)	4.0	4.0	2.8	3.0	6%	0.228
AutoDock4	0.0	0.0	1.6	1.6	0%	0.133
UCSF DOCK6	0.0	0.0	0.8	1.4	0%	0.073
COX-2						
GalaxyDock2 (flex.)	6.0	6.0	5.6	4.0	6%	0.336
GalaxyDock2 (rigid)	6.0	5.0	4.4	3.0	10%	0.277
AutoDock4	0.0	2.0	2.0	2.8	0%	0.180
UCSF DOCK6	0.0	0.0	0.0	0.2	0%	0.008
FXa						
GalaxyDock2 (flex.)	16.0	14.0	10.0	6.8	46%	0.649
GalaxyDock2 (rigid)	14.0	13.0	9.2	6.4	32%	0.589
AutoDock4	14.0	12.0	10.8	8.0	30%	0.677
UCSF DOCK6	16.0	11.0	7.2	4.8	22%	0.486
PDE5						
GalaxyDock2 (flex.)	8.0	5.0	5.8	3.8	10%	0.344
GalaxyDock2 (rigid)	6.0	4.0	4.0	3.2	8%	0.283
AutoDock4	8.0	5.0	2.8	3.0	10%	0.263
UCSF DOCK6	0.0	0.0	0.8	0.6	0%	0.057

Figure 4.4. Percentage of known inhibitors as a function of database coverage for GalaxyDock2 run in flexible mode and rigid mode, AutoDock4, and UCSF DOCK6 when virtual screening is performed on the Gilson set (Jorrisen and Gilson, 2005).



4.4. Conclusion of this section

GalaxyDock has been improved binding mode prediction accuracy by using a geometry-based method for generating the initial bank for CSA. With this method, the use of a smaller bank size has been made possible, increasing computational speed and at the same time enhancing the binding mode prediction performance. The enhancement in the performance of flexible side-chain docking is rather mild compared to that of rigid-receptor docking. Binding affinity prediction has also been improved by optimizing correlation with experimental binding affinity and root-mean-square error of predicted binding affinity by three-fold cross validation.

With its advanced binding mode and binding affinity prediction capability, GalaxyDock2 was applied to virtual screening. The program run in the rigid-receptor mode could produce the highest EF and BEDROC values on four tested proteins, when compared to AutoDock4 and UCSF DOCK6. Based on these results, it is concluded that GalaxyDock2 may be used as a new virtual screening tool. There still remain issues that have to be considered in the future such as further increasing the computation speed and improving the energy function. When receptor side-chain flexibility is considered in the virtual screening test, EF and BEDROC further increased in three of four proteins. The least successful example, CDK2, may have to be treated with a method that considers backbone flexibility of receptor explicitly. The next step of the GalaxyDock project is developing such a method

using the recently developed loop modeling techniques (Park and Seok, 2012, Ko *et al.*, 2011).

5. Conclusion of this thesis

Protein-ligand docking is a modeling technique that predicts binding conformation of protein and ligand and its binding affinity. Therefore, two components, sampling and scoring are improtant problems as in other modeling problems. In this thesis, three versions of program, LigDockCSA, GalaxyDock, and GalaxyDock2 are introduced. They incorporates CSA as a sampling method and appropriate scoring functions mainly based on AutoDock3 scoring function.

LigDockCSA, a first program, adds torsion part of PLP scoring function to AutoDock3 scoring function. In latter scoring function, there is a penalty that is concerned with conformational entropy being proportional to number of free torsion angles. However, it lacks torsion angle penalty that can be usually seen in other force fields. With AutoDock3 scoring function, CSA can find lower energy conformation than LGA but it shows higher deviations from native structure. Adding PLP torsion term modifies energy landscape and it helps to predict binding mode more accurately. The performance gap between CSA and LGA increases as number of free torsion angles increses. That is, CSA helps to find more accurate solution when degrees of freedom is high compared to other optimization technique.

Based on the fact, GalaxyDock, an extended version of LigDockCSA to flexible docking has been developed. Flexible receptor

docking is an important issue of PLD field because lots of failures of PLD applications come from protein flexibility. GalaxyDock expands LigDockCSA simply by adding degrees of freedom of side-chains and operators of exchanging side-chain χ angles. Energy function of the docking program should be modified because energy of protein can be changed due to its conformational variation. To find correct side-chain conformation, ROTA score and Lennard-Jones potential are added. With those components, GalaxyDock could find not only correct binding mode of ligand but also accurate side-chain conformation that is not reported and studied in other flexible docking papers.

GalaxyDock2 adds two components to its original version: initial bank generation using beta-complex and binding affinity prediction. Modification of initial bank by beta-complex strongly increases the its quality and this can help to predict binding mode more accurately. With the modification, GalaxyDock can reduce the size of bank from 100 to 30. Therefore, time consumption is also reduced. For binding affinity prediction, free ligand correction is introduced. It increases correlation between scoring function and experimentally observed binding affinity as in Huey's work. With those two components, based on the virtual screening benchmark result, GalaxyDock2 could be seemed as a new virtual screening tool.

Even though those programs shows a successful application of CSA to docking problem, there is a large room for improvements. The flexible side-chains should be designated before docking run. However, in real application world, the users would not know which residue

should be selected to be flexible. In addition, not only side-chains but also backbone conformation can be varied when ligand binds as shown in cAPK and CDK2 benchmark. Therefore, extension of flexible regions to all side-chains in binding pocket and backbone should be included. This increasement brings an explosion of degrees of freedom, however, strong aspects of CSA would solve the problem efficiently.

Binding affinity prediction needs to be improved also. Even free ligand correction increases the correlation between predicted binding affinity and experimental one, it is still lower than some re-socirng functions such as X-score and lower than 0.5. Implementation of other useful terms from other scoring function can helps to increase the precise prediction of binding affinity and it also helps to find binding mode more correctly because sampling and scoring are closely related. Two problems that mentioned above, are current issues in protein-ligand docking field. I hope that those issues will be improved by GalaxyDock in the future.

Appendix

Supplementary tables

Table S1. List of the 42 training set complexes that are selected randomly from the ASTEX diverse set for energy parameter determination

PDB ID	N_{tor}[a]	PDB ID	N_{tor}[a]	PDB ID	N_{tor}[a]
1q41	1	1p62	5	1z95	8
1hnn	1	1v4s	5	1n46	9
1p2y	1	1x8x	5	1s19	9
1uou	2	1hp0	6	2bsm	9
1gpk	3	1opk	6	1meh	10
1hww	3	1q1g	6	1sq5	10
1sg0	3	1t9b	6	2br1	10
1ia1	4	1lpz	7	1uml	11
1l2s	4	1u1c	7	1hvy	12
1oyt	4	1xoq	7	1mzc	12
1r9o	4	2bm2	7	1r58	12
1yvf	4	1m2z	8	1unl	12
1ke5	5	1sj0	8	1ygc	13
1of1	5	1xm6	8	1gkc	16

a. The number of rotatable bonds in the ligand

Table S2. RMSD of the docking pose predicted by AutoDock and LigDockCSA

PDB ID	N_{tor}	Auto Dock	LigDock CSA	PDB ID	N_{tor}	Auto Dock	LigDock CSA
1w1p	0	7.774	0.461	1lpz	7	0.702	0.757
1q41	1	0.799	0.540	1n2j	7	6.706	6.829
1u4d	1	0.937	0.793	1u1c	7	1.311	1.241
1hnn	1	1.102	0.918	1v48	7	0.633	0.817
1lrh	1	0.880	0.838	1xoq	7	1.365	1.345
1p2y	1	1.470	1.412	1yqy	7	1.044	0.963
1sqn	2	0.776	0.762	2bm2	7	1.145	1.008
1uou	2	0.962	0.874	1ig3	8	4.900	0.939
1xoz	2	0.651	0.576	1m2z	8	0.687	0.525
1gpk	3	1.293	1.397	1nav	8	0.367	0.401
1hq2	3	0.496	0.593	1sj0	8	0.965	0.754
1hww	3	0.401	0.376	1tz8	8	0.897	0.862
1owe	3	1.509	0.391	1xm6	8	1.762	2.145
1sg0	3	0.952	0.357	1ywr	8	0.855	0.780
1yv3	3	0.819	0.590	1z95	8	0.708	0.786
1ia1	4	0.833	0.718	1jla	9	0.985	0.782
1jd0	4	3.852	4.497	1n46	9	0.730	0.299
1l2s	4	2.862	5.590	1pmn	9	1.056	0.995
1n2v	4	2.946	2.683	1s19	9	0.777	0.707
1oyt	4	0.668	0.563	1t46	9	1.372	1.121
1q4g	4	0.569	0.550	2bsm	9	1.027	0.759
1r9o	4	5.552	1.374	1g9v	10	5.049	2.317
1tow	4	4.327	1.043	1meh	10	1.879	2.057
1yvf	4	0.689	0.808	1mmv	10	0.774	1.553
1j3j	5	0.793	0.481	1sq5	10	2.485	2.964
1ke5	5	0.644	0.579	1y6b	10	1.945	1.892
1n1m	5	0.798	0.569	2br1	10	3.333	1.527
1ofl	5	1.016	0.915	1l7f	11	0.889	0.676
1of6	5	0.676	0.595	1uml	11	1.222	1.670
1p62	5	0.520	0.410	1v0p	11	1.117	1.279
1tt1	5	0.839	0.736	1hvy	12	1.872	1.970
1v4s	5	1.427	1.481	1hwi	12	1.185	1.122
1w2g	5	0.464	0.670	1mzc	12	5.934	0.789
1x8x	5	2.506	2.936	1r1h	12	1.101	1.170
1gm8	6	2.367	8.016	1r58	12	3.426	2.929
1hp0	6	0.606	0.437	1s3v	12	4.804	4.811
1k3u	6	1.082	1.096	1unl	12	1.302	1.249
1opk	6	1.467	1.631	1vcj	12	1.147	0.745
1oq5	6	1.033	1.048	1ygc	13	1.278	2.679
1q1g	6	0.451	0.453	1kzk	15	0.801	0.759
1t40	6	0.488	0.555	1gkc	16	1.211	1.222
1t9b	6	1.237	1.148	1r55	16	0.877	0.913
1jje	7	0.796	0.794				

Table S3. Cross-docking results on the HIV protease set for (a) rigid docking using LigDockCSA, (b) GalaxyDock with 2 flexible residues (ARG8 of the two chains), and (c) GalaxyDock with 4 flexible residues (ARG8 and ILE50 of the two chains). Ligand RMSD of the predicted binding pose (starting from ligand A + protein structure B) from that of the crystal structure (ligand A + protein structure A) is presented for each cross docking of ligand A (row) to protein structure B (column).

(a)

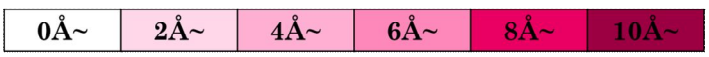
Ligands

Proteins	1HBV	1HEF	1HEG	1HIH	1HIV	1HPS	1HTE	1HTF	1HTG	1HVI	1HVJ	1HVK	1HVL	1HVS	1SBG	4HVP	4PHV	5HVP	8HVP	9HVP
1HBV	1.023	0.558	11.416	1.024	1.441	0.629	7.524	6.745	2.380	1.752	1.276	1.785	1.631	12.026	0.942	2.010	0.608	1.131	1.533	12.585
1HEF	1.708	1.508	1.436	0.674	0.929	1.145	9.938	3.269	0.910	1.594	12.102	14.770	1.669	12.073	0.739	1.111	10.758	1.087	1.538	0.588
1HEG	1.547	0.837	1.290	1.107	1.205	0.848	6.449	3.172	0.981	2.339	2.078	14.380	1.779	2.771	0.689	1.856	0.870	1.090	1.455	12.611
1HIH	10.954	1.386	11.393	0.698	1.551	0.826	6.594	2.797	1.129	2.213	2.032	12.099	2.299	12.907	9.758	2.371	0.767	1.647	1.435	1.147
1HIV	10.987	2.047	11.464	0.910	0.653	0.667	6.866	1.337	1.206	1.107	1.362	0.720	1.429	0.768	0.391	1.350	0.459	1.235	1.698	1.479
1HPS	1.739	1.209	1.349	0.890	0.954	0.556	10.011	1.004	1.267	2.133	1.935	14.877	2.082	1.187	1.450	1.474	0.627	1.135	1.356	1.062
1HTE	2.177	1.092	1.494	0.491	0.860	1.085	0.878	2.264	2.769	0.951	1.351	1.090	1.206	0.484	0.657	1.374	0.590	1.213	1.758	1.265
1HTF	11.026	1.026	1.388	1.046	1.270	0.467	6.133	1.096	0.415	11.232	11.294	1.739	2.118	1.823	0.731	2.341	0.555	1.222	1.416	1.070
1HTG	1.744	1.008	1.121	0.764	0.881	0.716	0.717	1.286	0.847	12.031	1.313	0.930	0.985	11.991	0.667	1.258	10.743	1.207	1.517	0.969
1HVI	10.985	1.575	11.409	1.365	1.496	1.381	6.771	1.580	2.822	1.063	1.016	1.320	1.135	0.830	1.209	1.861	0.673	1.111	1.447	1.499
1HVJ	1.665	1.632	1.298	1.226	1.508	1.452	7.042	1.522	2.415	1.018	0.959	1.205	0.899	0.762	1.488	1.185	0.635	1.214	1.970	1.601
1HVK	1.787	11.640	1.318	0.880	1.390	1.364	6.920	1.498	2.489	1.037	1.074	1.172	1.248	0.539	1.217	1.208	0.610	1.305	1.644	1.553
1HVL	1.812	1.578	1.380	1.294	1.753	1.486	6.914	1.562	2.451	1.024	1.074	1.138	0.922	1.065	1.283	1.293	10.757	1.307	1.504	1.554
1HVS	2.086	1.588	1.599	0.932	1.181	1.441	10.035	1.491	2.733	0.773	1.090	1.197	1.086	0.744	0.971	2.011	0.834	1.250	1.483	1.259
1SBG	11.095	1.311	1.120	0.951	1.103	0.484	8.296	3.177	0.995	1.722	1.602	14.908	1.450	2.246	0.827	2.352	0.643	0.976	1.406	12.605
4HVP	1.896	1.441	1.081	0.944	0.566	0.998	6.600	1.594	0.945	0.881	1.803	1.124	1.056	12.009	0.425	1.151	0.496	1.316	2.052	1.143
4PHV	10.983	11.685	1.270	0.789	1.206	1.378	6.363	2.663	0.987	2.429	1.808	14.102	11.866	1.384	1.062	2.291	0.875	0.854	1.265	0.804
5HVP	2.036	1.562	1.113	9.595	1.095	0.857	6.757	9.415	1.311	0.983	1.426	1.174	1.407	0.828	9.677	1.390	10.770	1.058	1.485	1.154
8HVP	10.893	2.009	1.958	1.227	1.018	0.774	1.262	1.784	2.541	1.245	1.160	12.086	1.376	0.833	9.747	1.383	0.532	0.821	1.563	1.186
9HVP	1.380	1.596	1.153	1.292	1.350	0.452	6.334	1.368	0.436	11.991	1.696	6.899	1.570	11.994	0.993	1.104	0.624	1.482	1.222	0.650

(b)

Ligands

Proteins	1HBV	1HEF	1HEG	1HIH	1HIV	1HPS	1HTE	1HTF	1HTG	1HVI	1HVJ	1HVK	1HVL	1HVS	1SBG	4HVP	4PHV	5HVP	8HVP	9HVP
1HBV	0.862	0.853	1.016	0.958	1.498	0.657	1.054	0.847	1.630	0.626	0.691	0.566	0.727	0.556	0.907	1.636	0.708	1.056	1.239	0.808
1HEF	1.406	1.115	1.156	0.649	1.788	0.846	0.829	0.630	0.763	1.126	0.669	0.570	0.570	1.262	1.780	1.566	0.592	0.949	1.261	0.525
1HEG	1.397	0.776	1.233	0.903	0.983	0.702	0.988	0.686	1.544	1.155	0.992	0.564	0.507	0.683	0.595	0.919	1.915	0.770	1.232	0.646
1HIH	1.347	1.269	1.169	4.581	1.047	0.506	0.874	0.874	1.589	0.629	0.621	0.674	0.789	0.277	1.103	1.112	0.600	1.021	1.134	0.971
1HIV	1.699	1.487	0.924	0.662	0.479	0.491	1.207	0.928	0.929	1.566	0.581	0.971	0.836	0.809	0.378	1.248	0.457	1.039	1.424	1.592
1HPS	1.257	1.676	1.111	0.791	0.784	0.526	0.967	0.675	1.694	0.671	0.875	0.638	0.652	0.763	1.194	1.349	0.568	0.868	1.152	0.718
1HTE	1.633	0.990	1.439	4.833	0.699	0.807	0.603	0.913	0.442	0.657	0.716	0.602	0.712	0.702	0.692	1.780	1.968	0.913	1.200	0.981
1HTF	1.575	0.697	1.077	0.903	0.989	0.736	0.543	0.783	1.596	0.716	0.845	0.645	0.764	0.800	1.466	1.345	0.626	1.003	1.175	0.863
1HTG	1.463	0.806	0.887	0.719	1.132	0.696	0.896	0.772	0.359	1.113	0.883	0.697	1.259	0.879	0.506	1.686	0.724	1.013	1.301	0.773
1HVI	1.414	1.083	1.132	0.822	1.477	1.083	1.319	1.279	1.759	0.862	0.666	0.686	0.643	0.507	1.884	1.624	0.566	0.966	1.139	1.156
1HVJ	1.330	1.096	1.015	0.843	1.186	1.001	1.272	1.203	1.809	0.669	0.597	0.710	0.720	0.474	0.982	0.996	0.591	1.028	1.355	1.185
1HVK	1.323	1.737	1.111	0.784	1.274	0.854	1.097	1.254	1.859	1.729	0.613	0.644	0.769	0.372	0.794	1.056	0.550	1.086	1.328	1.526
1HVL	1.399	1.040	1.135	0.874	1.315	0.901	1.317	1.067	1.772	0.588	0.535	0.735	0.708	0.526	1.009	1.051	0.583	1.012	1.157	1.596
1HVS	1.565	5.002	1.303	0.971	1.025	0.877	1.133	1.082	1.764	0.517	0.585	0.687	0.675	0.366	1.726	1.661	0.643	1.080	1.240	0.994
1SBG	1.316	1.105	0.974	0.417	0.982	0.390	0.790	0.768	1.556	0.863	1.049	0.608	0.611	0.674	0.665	1.619	0.688	0.791	1.229	1.460
4HVP	1.553	1.184	1.038	0.707	0.445	0.727	1.311	1.122	0.560	0.862	0.891	0.617	1.126	0.687	0.406	1.078	0.464	1.106	1.182	0.781
4PHV	1.586	1.590	1.108	0.784	0.921	0.397	0.719	0.986	1.544	0.898	0.914	0.708	0.904	0.958	0.811	0.941	0.494	0.792	1.098	1.522
5HVP	1.633	1.102	0.935	0.915	4.894	0.478	1.413	1.174	0.790	0.924	0.855	0.865	0.816	0.919	0.285	1.331	0.397	0.855	1.150	0.831
8HVP	1.626	1.584	0.897	0.914	0.903	0.818	1.085	1.256	1.531	0.608	0.779	0.647	0.874	0.578	4.494	0.925	0.436	0.864	1.273	1.448
9HVP	0.926	1.329	1.085	0.842	1.239	0.414	0.769	0.774	0.436	1.616	0.878	0.670	0.801	0.623	1.922	1.672	0.570	0.736	1.136	0.485



(c)

Ligands

Proteins	1HBV	1HEF	1HEG	1HIH	1HIV	1HPS	1HTE	1HTF	1HTG	1HVI	1HVJ	1HVK	1HVL	1HVS	1SBG	4HVP	4PHV	5HVP	8HVP	9HVP
1HBV	0.954	0.726	0.963	0.751	3.212	0.985	0.975	0.717	1.736	0.541	0.721	0.646	0.824	0.421	0.910	0.893	0.729	1.032	1.255	0.804
1HEF	1.393	1.734	1.002	4.910	0.608	0.511	0.851	0.795	1.680	1.494	1.694	0.568	1.042	1.025	0.631	1.658	0.652	1.048	1.246	0.445
1HEG	1.329	0.813	1.023	0.703	1.022	0.732	0.661	0.830	0.426	0.952	0.840	0.641	0.867	0.717	0.625	0.873	1.954	0.766	1.208	0.482
1HIH	1.380	1.262	1.147	2.834	1.205	0.646	0.702	0.739	0.928	0.684	0.614	0.663	0.721	0.417	1.829	1.737	0.595	0.974	1.122	1.001
1HIV	1.695	1.449	0.953	0.741	0.487	0.524	0.788	1.102	0.901	0.657	1.169	0.639	0.829	0.777	0.492	1.362	0.464	1.085	1.448	1.134
1HPS	1.323	0.751	1.098	0.741	0.819	0.494	1.001	0.718	0.638	0.670	1.622	0.620	0.797	0.984	1.238	0.922	0.543	0.848	1.093	1.382
1HTE	1.887	0.843	1.218	4.938	0.688	0.754	0.885	0.930	1.737	1.429	0.748	0.587	0.681	0.704	0.626	1.173	0.813	0.848	1.178	0.890
1HTF	1.438	0.800	1.110	0.649	0.904	0.475	0.256	0.762	1.622	1.069	0.888	0.795	0.756	0.797	0.441	1.031	0.674	0.983	1.173	0.833
1HTG	1.632	0.654	1.054	0.511	0.718	0.639	0.585	0.696	0.498	1.290	0.878	0.699	0.962	0.892	0.607	1.790	0.571	1.014	1.311	0.801
1HVI	1.292	1.238	1.161	0.935	1.491	0.614	1.113	1.153	1.803	0.881	0.674	0.795	0.800	0.745	0.678	1.616	0.568	1.034	1.142	1.150
1HVJ	1.442	1.236	1.010	0.642	1.194	0.890	1.241	1.135	1.773	0.642	0.734	1.369	1.418	0.496	0.902	0.984	0.556	1.018	1.459	1.228
1HVK	1.389	1.591	1.122	0.591	1.423	0.842	4.604	1.144	1.727	0.755	0.593	0.664	0.720	0.507	0.855	1.071	0.538	1.114	1.349	1.171
1HVL	1.419	1.079	1.092	0.788	1.217	0.877	1.413	1.227	1.774	0.700	0.727	0.726	0.795	0.554	1.860	1.719	0.548	1.076	1.188	1.222
1HVS	1.657	1.264	1.408	0.666	1.152	0.632	1.090	1.127	1.712	0.667	0.957	0.631	0.660	0.435	0.700	1.522	0.601	1.081	1.232	1.106
1SBG	1.308	1.166	0.989	0.635	0.966	0.461	0.484	0.815	1.553	0.837	0.841	0.536	0.658	0.610	0.500	1.631	0.633	0.784	1.140	0.597
4HVP	1.542	1.506	0.692	0.694	0.458	0.709	4.563	1.092	0.604	0.950	1.039	1.071	0.676	1.056	0.452	1.245	0.402	1.079	1.387	0.652
4PHV	1.280	0.831	1.091	0.651	1.021	0.338	0.864	0.847	1.583	1.011	0.829	0.686	1.639	1.014	0.798	0.959	0.644	0.779	1.110	0.650
5HVP	1.586	0.962	0.889	0.804	0.973	0.416	1.353	1.200	0.826	1.564	0.847	1.080	0.801	0.917	1.649	1.240	0.512	0.744	1.265	1.293
8HVP	1.411	1.603	1.103	1.374	0.877	0.693	1.162	1.240	1.547	0.899	0.851	0.936	0.835	0.546	4.716	0.933	0.424	0.827	1.274	0.976
9HVP	1.157	1.590	1.003	0.339	1.095	2.942	0.638	0.650	0.471	0.735	0.811	0.671	0.690	0.642	0.964	0.563	0.566	0.724	1.098	0.467

0Å~	2Å~	4Å~	6Å~	8Å~	10Å~

Table S4. Accuracy of χ_1 angles for flexible docking with GalaxyDock for (a) 2 flexible residues (2 ARG8's) and (b) 4 flexible residues (2 ARG8's and 2 ILE50's). The χ_1 angle prediction is considered accurate is its value is within 30° from the native value.

(a)

Ligands

Proteins	1HBV	1HEF	1HEG	1HIH	1HIV	1HPS	1HTE	1HTF	1HTG	1HVI	1HVJ	1HVK	1HVL	1HVS	1SBG	4HVP	4PHV	5HVP	8HVP	9HVP
1HBV	1.000	0.500	0.500	1.000	1.000	1.000	0.500	1.000	1.000	1.000	0.500	0.500	0.000	0.500	1.000	0.500	1.000	1.000	1.000	0.000
1HEF	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.500
1HEG	1.000	0.500	0.000	0.000	0.000	0.500	1.000	0.000	0.000	1.000	1.000	0.000	1.000	0.500	0.500	1.000	0.000	0.500	0.000	1.000
1HIH	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	0.000
1HIV	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
1HPS	0.500	1.000	0.500	1.000	1.000	1.000	0.500	0.500	0.500	0.500	1.000	0.500	0.500	0.500	1.000	1.000	1.000	0.500	1.000	0.500
1HTE	0.500	0.500	0.500	0.500	0.500	0.500	1.000	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
1HTF	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	0.500
1HTG	1.000	1.000	1.000	1.000	1.000	0.500	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	0.500
1HVI	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
1HVJ	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
1HVK	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
1HVL	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
1HVS	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.500
1SBG	1.000	1.000	1.000	0.500	1.000	1.000	0.500	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.500
4HVP	1.000	1.000	1.000	1.000	1.000	1.000	0.500	0.500	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000	0.500
4PHV	1.000	1.000	1.000	1.000	0.500	1.000	0.500	1.000	1.000	1.000	0.500	0.000	0.500	0.500	1.000	1.000	1.000	1.000	1.000	0.000
5HVP	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.500
8IVP	1.000	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
9HVP	0.500	1.000	0.500	1.000	1.000	0.500	0.500	0.500	1.000	1.000	0.500	1.000	1.000	0.500	1.000	1.000	0.500	1.000	1.000	0.500

100%	50%	0%

(b)

Ligands

Proteins	1HBV	1HEF	1HEG	1HIH	1HIV	1HPS	1HTE	1HTF	1HTG	1HVI	1HVJ	1HVK	1HVL	1HVS	1SBG	4HVP	4PHV	5HVP	8HVP	9HVP
1HBV	1.000	1.000	1.000	0.500	0.750	0.750	0.250	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.750	0.750	0.500	0.500	0.500	0.500
1HEF	0.500	0.750	0.500	0.750	0.500	0.500	0.250	0.500	0.500	0.750	0.750	0.500	0.750	0.500	0.500	0.500	0.500	0.500	0.500	0.250
1HEG	0.500	0.750	1.000	0.250	0.250	0.750	0.250	0.000	0.500	0.250	0.250	0.250	0.500	0.250	0.500	0.750	0.250	0.250	0.500	0.750
1HIH	0.500	0.500	0.500	1.000	0.750	0.500	0.500	0.750	1.000	1.000	0.750	1.000	0.750	1.000	0.500	0.750	0.750	0.750	0.500	0.000
1HIV	0.500	0.500	0.500	0.750	0.750	0.500	0.500	0.500	0.750	0.500	0.750	0.750	0.750	0.500	0.500	0.750	0.750	0.500	0.750	0.000
1HPS	0.750	0.750	0.750	0.500	0.750	0.750	0.250	0.500	0.500	0.250	0.500	0.500	0.500	0.250	0.500	0.750	0.250	0.500	1.000	0.500
1HTE	0.000	0.250	0.000	0.250	0.500	0.500	1.000	0.750	0.500	0.500	0.500	0.500	0.500	0.750	0.250	0.250	0.500	0.750	0.000	0.250
1HTF	0.500	0.500	0.500	1.000	0.750	0.500	0.500	1.000	0.750	1.000	1.000	0.750	0.750	1.000	0.500	0.750	0.750	0.500	0.500	0.000
1HTG	0.500	0.500	0.500	1.000	0.750	0.500	0.750	1.000	1.000	1.000	0.750	0.750	0.500	0.750	0.500	0.500	1.000	1.000	0.500	0.250
1HVI	0.500	0.500	0.500	1.000	0.500	0.500	0.500	1.000	1.000	1.000	0.750	0.750	0.750	1.000	0.500	0.500	1.000	1.000	0.750	0.000
1HVJ	0.500	0.500	0.500	0.750	0.750	0.750	0.250	0.750	0.750	0.750	1.000	0.750	1.000	0.750	0.750	0.750	0.750	0.750	0.750	0.000
1HVK	0.500	0.500	0.500	1.000	0.750	0.750	0.250	0.750	1.000	1.000	1.000	0.750	1.000	0.750	0.750	0.750	0.750	0.750	0.750	0.250
1HVL	0.500	0.500	0.500	0.750	0.750	0.500	0.250	0.750	0.750	0.750	1.000	1.000	1.000	0.750	0.750	0.750	0.750	0.750	0.750	0.000
1HVS	0.500	0.500	0.500	1.000	0.750	0.500	0.250	0.750	1.000	1.000	0.750	0.750	0.750	1.000	0.750	0.750	0.750	1.000	0.750	0.250
1SBG	0.750	0.750	0.750	0.500	1.000	1.000	0.250	0.500	0.500	0.250	0.750	0.750	0.750	0.500	1.000	1.000	0.500	0.500	1.000	0.500
4HVP	0.500	0.500	0.500	0.750	0.750	0.500	0.250	0.500	0.500	0.250	0.750	0.750	0.500	0.500	0.500	0.750	0.750	0.750	0.750	0.000
4PHV	0.500	0.500	0.500	0.750	0.500	0.500	0.500	1.000	0.750	0.750	0.750	1.000	1.000	1.000	0.750	0.500	0.750	1.000	0.500	0.000
5HVP	0.500	0.500	0.500	1.000	0.500	0.500	0.750	1.000	1.000	1.000	0.500	0.750	0.750	0.750	0.500	0.500	0.750	1.000	0.500	0.250
8IVP	0.500	0.500	0.500	0.500	0.750	0.500	0.500	0.500	0.500	0.500	0.750	0.750	1.000	0.500	0.750	0.750	0.500	0.500	0.750	0.000
9HVP	0.500	1.000	1.000	0.500	0.750	0.750	0.500	0.500	0.500	0.250	0.500	0.500	0.750	0.250	0.750	0.750	0.500	0.500	0.750	0.500

100%	75%	50%	25%	0%

Table S5. Comparison of the LXRβ cross-docking experiment with RosettaLigand (Davis and Baker, 2009) and AutoDock4.

Receptor	Ligand	Galaxy Dock	Rosetta Ligand 1[a]	Rosetta Ligand 2[b]	Auto Dock4[c]	LigDock CSA
1P8D	1P8D	1.33 Å	1.49 Å	1.39 Å	2.11 Å	2.01 Å
	1PQ6	2.41 Å	2.06 Å	2.24 Å	4.82 Å	3.73 Å
	1PQC	1.97 Å	1.78 Å	2.29 Å	4.42 Å	4.58 Å
1PQ6	1P8D	1.75 Å	1.20 Å	1.27 Å	2.91 Å	2.08 Å
	1PQ6	0.89 Å	2.05 Å	2.08 Å	1.45 Å	1.64 Å
	1PQC	1.84 Å	3.46 Å	3.61 Å	3.65 Å	2.41 Å
1PQC	1P8D	1.60 Å	0.98 Å	0.98 Å	4.57 Å	9.27 Å
	1PQ6	1.90 Å	3.70 Å	2.20 Å	2.73 Å	4.31 Å
	1PQC	1.00 Å	0.49 Å	0.53 Å	1.62 Å	1.81 Å
Average		1.63 Å	1.91 Å	1.84 Å	3.14 Å	3.64 Å
Success Rate		88.9%	55.6%	44.4%	22.2%	22.2%

a. Flexible side-chain with rigid backbone.

b. Backbone flexibility allowed during minimization after side-chain and ligand placing.

c. Rigid docking. The values were taken from the RosettaLigand paper (Davis and Baker, 2009)

Table S6. Cross-docking results of GalaxyDock for the cAPK set. RMSD (in Å) and χ_1 angle accuracy (in %, in parenthesis) are shown.

Receptor \ Ligand	1BKX	1BX6	1STC	1YDT
1BKX	1.11 (100)	2.94 (100)	2.93 (100)	1.37 (100)
1BX6	0.99 (50)	2.96 (100)	4.15 (50)	0.96 (50)
1STC	1.04 (0)	3.75 (50)	0.56 (50)	5.92 (50)
1YDT	1.06 (100)	1.19 (100)	0.97 (100)	1.21 (100)

Table S7. Cross-docking results of GalaxyDock for the diverse set. RMSD (in Å) is shown.

Protein	Receptor	Ligand	Experimental pocket		Predicted pocket	
			Galaxy Dock	LigDock CSA	Galaxy Dock	LigDock CSA
Aldose Reductase	2ACR	2FZB	3.651	4.882	3.430	3.653
Anti-Steroid Fab	1DBA	1DBB	4.318	6.568	3.825	6.822
CDK2	1AQ1	1DM2	1.742	6.628	1.206	1.764
	1DM2	1AQ1	0.512	0.788	1.314	1.124
COX-2	1CX2	3PGH	0.986	1.042	1.010	0.952
	3PGH	1CX2	1.055	1.007	0.829	1.088
Estrogen Receptor	1ERR	3ERT	1.646	1.239	1.894	2.435
	3ERT	1ERR	1.297	2.518	1.563	1.203
Factor Xa	1KSN	1XKA	1.687	1.585	1.686	2.941
	1XKA	1KSN	0.797	2.485	4.184	4.892
GSK-3 β	1Q4L	1UV5	0.825	1.219	0.956	1.219
	1UV5	1Q4L	1.945	2.637	1.974	2.637
Hiv1 RT	1C1C	1RTH	5.273	3.310	3.534	2.908
	1RTH	1C1C	1.482	5.933	1.450	1.368
JNK3	1PMN	1PMV	1.087	1.116	0.919	5.071
	1PMV	1PMN	0.911	5.549	0.920	5.058
LXR β	1P8D	1PQ6	1.955	2.113	1.888	3.274
	1PQ6	1P8D	1.637	1.798	1.749	1.734
Neuroaminidase	1A4Q	1NSC	0.911	1.107	1.135	1.181
	1NSC	1A4Q	1.009	4.476	1.387	2.288
P38 Kinase	1BMK	1DI9	1.758	5.980	3.635	4.006
	1DI9	1BMK	1.197	1.025	1.048	4.023
PKA	1STC	1YDS	1.601	1.283	1.568	1.310
	1YDS	1STC	1.737	3.914	1.728	2.096
PPARγ	1FM9	2PRG	1.296	5.255	5.655	3.135
	2PRG	1FM9	4.459	3.830	6.394	7.891
TK	1KI4	1KIM	1.419	1.398	1.462	0.870
	1KIM	1KI4	0.894	4.120	0.907	1.367
Trypsin	1PPC	1PPH	1.069	8.458	0.680	2.904
	1PPH	1PPC	1.691	1.655	1.700	1.655

Table S8. Performance dependence of GalaxyDock1 on the beta value and the offset distance when tested on the ASTEX diverse set.

(a) Dependence on the beta value (with offset distance set to the van der Waals radius of the largest ligand atom)

Beta value (Å)	1.0	1.2	1.4	1.6	1.8
Success Rate[a] (%)	76.5	81.2	72.9	76.5	76.5
Average RMSD (Å)	1.58	1.44	1.69	1.61	1.53

a. Percentage of the cases in which RMSD of the top scoring pose from the experimental pose is < 2 Å.

(b) Dependence on the offset distance (with beta value = 1.2 Å)

Offset distance (Å)	1.5	1.6	1.7	1.8	1.9	2.0
Success Rate (%)	74.1	78.8	80.0	84.7	80.0	75.3
Average RMSD (Å)	1.78	1.56	1.64	1.29	1.53	1.69

Bibliography

Beautrait, A. *et al.* (2008) Induced fit in Liver X receptor beta: A molecular dynamics-based investigation. *Proteins: Struct. Funct. Bioinf.* 72, 873-882.

Block, P. *et al.* (2006) AffinDB: a freely accessible database of affinities for protein–ligand complexes from the PDB. *Nucleic Acids Res.* 34, D522-D526.

Bottegoni, G., Kufareva, I. *et al.* (2008) A new method for ligand docking to flexible receptors by dual alanine scanning and refinement. *J. Comput. -Aided Mol. Des.* 22, 311-325.

Bottegoni, G. *et al.* (2009) Four-Dimensional Docking: A Fast and Accurate Account of Discrete Receptor Flexibility in Ligand Docking. *J. Med. Chem.* 52, 397-406.

Brik, A. and Wong, C. -H. (2003) HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* 1, 5-14.

Cavasotto, C. N., Kovacs, J. A. *et al.* (2005) Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes. *J. Am. Chem. Soc.* 127, 9632-9640.

Chen, H. -M., Liu, B. -F. *et al.* (2007) SODOCK: Swarm Optimization for Highly Flexible Protein-Ligand Docking. *J. Comput. Chem.* 28, 612-623.

Cheng, T. *et al.* (2009) Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Info. Model.* 49, 1079-1093.

Clauβen, H., Buning, C. *et al.* (2001) FlexE: Efficient Molecular Docking Considering Protein Structure Variation. *J. Mol. Biol.* 308, 377-395.

Connolly, M. L. (1993) The molecular surface package. *J. Mol. Graphics* 11, 139-141.

Corbeil, C. R., Englebienne, P. *et al.* (2007) Docking Ligands into Flexible and Solvated Macromolecules. 1. Development and Validation of FITTED 1.0. *J. Chem. Inf. Model.* 47, 435-449.

Davis, I. W. and Baker, D. (2009) ROSETTALIGAND Docking with Full Ligand and Receptor Flexibility. *J. Mol. Biol.* 385, 381-392.

Dunbrack, R. L. and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* 6, 1661-1681.

Dunbrack, R. L. (2002) Rotamer Libraries in the 21st Century. *Curr. Opin. Str. Biol.* 12, 431-440.

Engh, R. A. *et al.* (1996) Crystal Structures of Catalytic Subunit of cAMP-dependent Protein Kinase in Complex with Isoquinolinesylfonyl Protein Kinase Inhibitors H7, H8, and H9. *J. Biol. Chem.* 271, 26157-26164.

Färnegårdh, M. *et al.* (2003) The Three-dimensional Structure of the Liver X Receptor β Reveals a Flexible Ligand-binding Pocket That Can Accomodate Fundamentally Different Ligands. *J. Biol. Chem.* 278, 38821-38828.

Gaudreault, F. *et al.* (2012) Side-chain rotamer changes upon ligand

binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. *Bioinformatics* 28, i423-i430.

Gehlharr, D. K. *et al.* (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* 2, 317-324.

Goodford, P. J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 7, 849-857.

Hartmann, C. e*t al.* (2007) IRECS: A new algorithm for the selection of the most probable ensembles of side-chain conformations in protein models. *Protein Sci.* 16, 1294-1307.

Hartmann, C. *et al.* (2009) Docking and scoring with alternative side-chain conformations. *Proteins: Struct. Funct. Bioinf.* 74, 712-726.

Hartshorn, M. J., Verdonk, M. L. *et al.* (2007) Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* 50, 726-741.

Heringa, J. *et al.* (1999) Strain in Protein Structure as Viewed Through Nonrotameric Side Chains: II. Effects Upon Ligand Binding. *Proteins: Struct. Funct. Bioinf.* 37, 44-55.

Huang, S. -Y., Grinter, S. Z. *et al.* (2010) Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* 12, 12899-12908.

Huang, S. -Y. and Zou, X. (2007) Ensemble Docking of Multiple

Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins: Struct. Funct. Bioinf.* 66, 399-421.

Huey, R. *et al.* (2007) A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28, 1145-1152.

Jackson, R. M. (1999) Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: Implications for the protein docking problem. *Protein Sci.* 8, 603-613.

Janson, S. *et al.* (2008) Molecular docking with multi-objective Particle Swarm Optimization. *Appl. Soft Comput.* 8, 666-675.

Jeffrey, P. D. *et al.* (1995) Mechanism of CDK2 activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* 376, 313-320.

Jones, G. *et al.* (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* 245, 43-53.

Jones, G., Willett, P. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727-748.

Joo, K., Lee, J. *et al.* (2009) All-atom chain-building by optimizing MODELLER energy function using conformational space annealing, *Proteins: Struct. Funct. Bioinf.* 75, 1010-1023.

Jorissen, R. N. and Gilson, M. K. (2005) Virtual Screening of Molecular Databases Using a Support Vector Machine. *J. Chem. Inf. Model.* 45, 549−561.

Kelley, L. A. *et al.* (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* 9 1063-1065.

Kim, D. -S. *et al.* (2010) Three-dimensional beta-shapes and beta-complexes via quasi-triangulation. *Comput. -Aided Des.* 42, 911-929.

Kim, D. -S. *et al.* (2011) BetaDock: Shape-Priority Docking Method Based on Beta-Complex. *J. Biomol. Struct. Dyn.* 29, 219-242.

Klon, A. E. *et al.* (2004) Finding More Needles in the Haystack: A Simple and Efficient Method for Improving High-Throughput Docking Results. *J. Med. Chem.* 47, 2743-2749.

Ko, J. *et al.* (2011) The FALC-Loop web server for protein loop modeling. *Nucleic Acids Res.* 39, W210-W214.

Kokh, D. B. and Wenzel, W. (2008) Flexible Side Chain Models Improve Enrichment Rates in In Silico Screening. *J. Med. Chem.* 51, 5919–5931.

Korb, O. *et al.* (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Info. Model.* 49, 84-96.

Kuntz, I. D., Blaney, J. M. e*t al.* (1982) A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 161, 269-288.

Lavecchia, A. e*t al.* (2006) Modeling of Cdc25B Dual Specificity Protein Phosphate Inhibitors: Docking of Ligands and Enzymatic Inhibition Mechanism. *ChemMedChem* 1, 540-550.

Lee, J. *et al.* (1997) New Optimization Method for Conformational

Energy Calculations on Polypeptides: Conformational Space Annealing. *J. Comput. Chem.* 18, 1222-1232.

Lee, J. and Seok, C. (2009) A statistical rescoring scheme for protein–ligand docking: Consideration of entropic effect. *Proteins: Struct. Funct. Bioinf.* 70, 1074-1083.

Lee, K. *et al.* (2005a) Study of Protein-Protein Interaction Using Conformational Space Annealing, *Proteins: Struct. Funct. Bioinf.* 60, 257-262.

Lee, K. *et al.* (2005b) An efficient molecular docking using conformational space annealing. *J. Comput. Chem.* 26, 78-87.

Lu, B. and Wong, C. F. (2005) Direct estimation of entropy loss due to reduced translational and rotational motions upon molecular binding. *Biopolymers* 79, 277-285.

Meng, E. C. *et al.* (1992) Automated docking with grid-based energy evaluation. *J. Comput. Chem.* 13, 505-524.

Morris, G. M., Goodsell, D. S. e*t al.* (1998) Automated docking using a Lamarkian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639-1662.

Morris, G. M. *et al.* (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* 2009, 30, 2785-2791.

Mukherjee, S., Balius, T. E. *et al.* (2010) Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *J. Chem. Inf. Model.* 50, 1986-2000.

Najmanovich, R. *et al.* (2000) Side-Chain Flexibility in Proteins upon

Ligand Binding. *Proteins: Struct. Funct. Bioinf.* 39, 261-268.

Neider, J. A. and Mead, R. (1965) A simplex method for function minimization. *The Comput. J.* 7, 308-313.

Österberg, F. *et al.* (2002) Automated Docking to Multiple Target Structures: Incorporation of Protein Mobility and Structural Water Heterogeneity in AutoDock. *Proteins: Struct. Funct. Bioinf.* 46, 34-40.

Park, B. J. *et al.* (2003) A hybrid genetic algorithm for the job shop scheduling problems. *Comput. Ind. Eng.* 45, 597-613.

Park, H. *et al.* (2011) Refinement of protein termini in template-based modleing using conformational space annealing. *Proteins: Struct. Funct. Bioinf.* 79, 2725-2734.

Park, H. and Seok, C. (2012) Refinement of unreliable local regions in template-based protein models, *Proteins: Struct. Funct. Bioinf.* 80, 1974-1986.

Patterson, E. F. *et al.* (2004) UCSF Chimera－A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605-1612.

Rarey, M., Lengauer, T. e*t al.* (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470-489.

Reeves, C. R. (1995) A GENETIC ALGORITHM FOR FLOWSHOP SEQUENCING. *Comput. Oper. Res.* 22, 5-13.

Rogers, J. P., Beuscher, A. E. e*t al.* (2006) Discovery of Protein Phosphatase 2C Inhibitors by Virtual Screening. *J. Med. Chem.*

49, 1658-1667.

Ruvinsky, A. M. Role of binding entropy in the refinement of protein‐ligand docking predictions: Analysis based on the use of 11 scoring functions. *J. Comput. Chem.* 28, 1364-1372.

Sherman, W., Day, T. *et al.* (2006) Novel Procedure for Modeling Ligand/Protein Induced Fit Effects. *J. Med. Chem.* 49, 534-553.

Schmidt, T. et al. (2011) Assessment of ligand-binding residue prediction in CASP 9. *Proteins: Struct. Funct. Bioinf.* 79 (Suppl 10), 126-136.

Shin, W. -H. *et al.* (2011) LigDockCSA: Protein-Ligand Docking Using Conformational Space Annealing. *J. Comput. Chem.* 32, 3226-3232.

Shin, W. -H. and Seok, C. (2012) GalaxyDock: Protein-Ligand Docking with Flexible Side-chains. *J. Chem. Info. Model.* 52, 3225-3232.

Snarey, M. *et al.* (1997) Comparison of algorithms for dissimilarity-based compound selection. *J. Mol. Graphics. Model.* 15, 372-385.

Sousa, S. F. *et al.* (2006) Protein-Ligand Docking: Current Status and Future Challenges. *Proteins: Struct. Funct. Bioinf.* 65, 15-26.

Spitzer, R. and Jain, A. N. (2012) Surflex-Dock: Docking benchmarks and real-world application. *J. Comput. -Aided Mol. Des.* 26, 687–699.

Steinland, K. *et al.* (2007) From modeling to Medicinal Chemistry: Automatic Generation of Two-Dimensional Complex Diagrams. *ChemMedChem* 2, 853-560.

Triballeau, N. *et al.* (2005) Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* 2005, 48, 2534-2547.

Truchon, C. F. *et al.* (2007) Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Info. Model.* 47, 488-508.

Velec, H. F. G., Gohlke, H. *et al.* (2002) DrugScore^{CSD}－Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *J. Med. Chem.* 48, 6296-6303.

Venkatesan, S. K. *et al.* (2010) Molecular Docking Studies of Selected Tricyclic and Quinone Derivatives on Trypanothione Reductase of Leishmania infantum. *J. Comput. Chem.* 31, 2463-2475.

Wang, R. *et al.* (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. -Aided Mol. Des.* 16, 11-26.

Wong, C. -F. *et al.* (2005) Molecular Docking of Balanol to Dynamics Snapshots of Protein Kinase A. *Proteins: Struct. Funct. Bioinf.* 61, 850-858.

Yanover, C. *et al.* (2008) Minimizing and Learning Energy Functions for Side-Chain Prediction. *J. Comput. Biol.* 15, 899-911.

Zavodsky, M. I. *et al.* (2005) Side-chain flexibility in protein-ligand

binding: The minimal rotation hypothesis. *Protein Sci.* 14, 1104-1114.

Zhao, Y. and Sanner, M. F. (2007) FLIPDock: Docking flexible ligands into flexible receptors. *Proteins: Struct. Funct. Bioinf.* 68, 726-737.

Zhao, Y. and Sanner, M. F. (2008) Protein-ligand docking with multiple flexible side-chains. *J. Comput. -Aided. Mol. Des.* 22, 673-679.

Zsolodos, Z., Reid, D. *et al.* (2007) eHiTS: a new fast, exhaustive flexible ligand docking system. *J. Mol. Graphics Modell.* 15, 411-428.

국문 초록

단백질-리간드 도킹은 1980년대에 처음 개발된 이래로 컴퓨터 기반 신약 개발의 중요한 도구가 되었다. 도킹의 목표는 주어진 단백질-리간드 복합체의 1) 결합구조 2) 결합 친화도를 정확하게 예측하는 것이다. 결합 구조의 정확한 예측은 단백질과 리간드의 구조의 적절한 샘플링을 필요로 한다. 많은 수의 도킹 프로그램들은 상대적으로 적은 수의 자유도가 요구되는 리간드의 샘플링은 성공적으로 수행한다. 그러나 현재 도킹 프로그램들의 많은 수가 리간드가 결합할 때, 수용체의 구조가 변함에도 불구하고 그 것의 구조적 유연성을 다루기가 쉽지 않기 때문에 단백질을 강체로 가정하여 계산을 수행한다. 무엇보다, 주로 사용되는 샘플링 방법으로는 수용체의 큰 구조 공간을 탐색하는 것이 적절히 이루어지지 않는다. 덧붙여서, 실험적 도킹 평가함수나 역장 기반 에너지 함수와 같은 현재의 에너지 함수들은 아직까지는 유연한 수용체와 유연한 리간드간의 상호작용을 정확히 기술하지 못한다.

본 논문에서는 수용체의 구조 유연성을 효과적으로 고려하는 도킹 프로그램인 GalaxyDock의 발전 과정이 기술되어 있다. 이 프로그램에서는 강력한 광역 최적화 기술인 conformational space annealing을 도입하여 단백질과 리간드의 구조 공간을 동시에 탐색한다. 덧붙여, 유연한 수용체 도킹에 사용되는 새로운 에너지 함수를 도입한다. 이 함수는 지식-기반 평가함수인 ROTA와 역장-기반 함수인 AutoDock 에너지 함수를 결합하여 만들어 진다. 샘플링과 평가함수의 이러한 요소들로써, GalaxyDock은 현재의 여러 도킹 프

로그램들과 비교해 보았을 때, 결합 구조 예측과 가상 검색 분야에 뛰어난 성능을 보여준다. 이 결과는 GalaxyDock이 더욱 발전된 방법의 기반이 될 수 있으며 in sillico 신약 개발 과정에 실용적인 방법이 될 수 있음을 보여준다.

핵심단어: 단백질-리간드 도킹, 광역 최적화, 가상 검색, 컴퓨터 기반 신약 개발
학번: 2008-20327